WEST Search History

Hide Items | Restore | Clear | Cancel

DATE: Wednesday, June 09, 2004

Hide?	Set Name	Query	Hit Count
		DB=USPT; PLUR=YES; OP=AND	
☐	L1	antiabc or anti-abc	2
☐	L2	antiabctransport$ or anti-abctransport$	0
☐	L3	abctransport$ or abc-transport$	25
☐	L4	L3 not l1	25
☐	L5	L4 and (antibody or antibodies or immune or globulin or igg or igm or iga or sfv or fab or chimeric or chimaeric or humanized or humanization or humanizing)	22
☐	L6	L4 and (antibody or monoclonal or mab or moab or mono-clonal or clonal or antibodies or immune or globulin or igg or igm or iga or sfv or fab or chimeric or chimaeric or humanized or humanization or humanizing)	22
☐	L7	L4 same (antibody or monoclonal or mab or moab or mono-clonal or clonal or antibodies or immune or globulin or igg or igm or iga or sfv or fab or chimeric or chimaeric or humanized or humanization or humanizing)	2
☐	L8	L3.clm. and (antibody or monoclonal or mab or moab or mono-clonal or clonal or antibodies or immune or globulin or igg or igm or iga or sfv or fab or chimeric or chimaeric or humanized or humanization or humanizing).clm.	0

END OF SEARCH HISTORY

WEST Search History

Hide Items | Restore | Clear | Cancel

DATE: Wednesday, June 09, 2004

Hide?	Set Name	Query	Hit Count
		DB=USPT; PLUR=YES; OP=AND	
☐	L1	atp near4 transport	315
☐	L2	L1 same (gram near4 positive)	7
		DB=PGPB,USPT,USOC,EPAB,JPAB,DWPI,TDBD; PLUR=YES; OP=AND	
☐	L3	abc.ti,ab,clm.	5947
☐	L4	L3 same (transport$ or system$2 or cassett$2 or complex$2 or operon$2).ti,ab,clm.	985
☐	L5	L4 and (strep$ or staph$ or (gram near3 positive))	73
		DB=USPT; PLUR=YES; OP=AND	
☐	L6	US-6251629-B1.did.	1

END OF SEARCH HISTORY

Generate Collection | Print

Search Results - Record(s) 51 through 73 of 73 returned.

☐ 51. 6077826. 08 Jun 98; 20 Jun 00. Synthetic macromolecular channel assembly for transport of chloride ions through epithelium useful in treating cystic fibrosis. Tomich; John M., et al. 514/12; 514/13 530/324 530/325 530/326. C07K014/00 A61K038/16.

☐ 52. 5773277. 18 May 95; 30 Jun 98. Crystalline chondroitinase isolated from Proteus vulgaris ATCC 6896. Hashimoto; Nobukazu, et al. 435/232; 435/183 435/188 435/201. C12N009/88 C12N009/44.

☐ 53. JP411253178A. 07 Oct 98. 21 Sep 99. ABC TRANSPORTER. WARREN, RICHARD L. C12N015/09; C07K014/31 C07K016/12 C12P021/02 C12Q001/68 G01N033/53 G01N033/577 A61K031/00 A61K038/00 A61K048/00.

☐ 54. WO003087147A1. 14 Apr 03. 23 Oct 03. STRETPTOCOCCAL GENES INVOLVED IN OSMOTIC AND OXIDATIVE STRESS AND IN VIRULENCE. BROWN, JEREMY STUART. C07K014/315; A61K035/74 C07K016/12.

☐ 55. WO009957281A2. 06 May 99. 11 Nov 99. PEPTIDE ANTIBIOTICS WHICH INHIBIT THE GROWTH OF PNEUMOCOCCI, ABC TRANSPORTER AND TWO-COMPONENT SIGNAL TRANSDUCTION SYSTEM PROTEINS FROM STREPTOCOCCUS PNEUMONIAE, AND METHODS USING THE SAME. NOVAK, RODGER, et al. C12N015/31; C12N015/54 C12N015/55 C12N009/12 C12N009/14 C07K007/08 C07K014/315 C07K016/12 C12Q001/68 A61K039/09.

☐ 56. WO009801154A2. 07 Jul 97. 15 Jan 98. TREATMENT AND DIAGNOSIS OF INFECTIONS OF GRAM POSITIVE COCCI. BURNIE, JAMES PETER, et al. A61K039/00; A61K039/02 A61K039/09 A61K039/40 G01N033/569 C07K014/31 C07K016/12 A61K039/085.

☐ 57. WO2003087147A. Peptides derived from phg ABC operon, useful for the manufacture of a medicament for treating or preventing a condition associated with infection by Streptococcus pneumoniae or other gram-positive bacteria. BROWN, J S. A61K035/74 C07K014/315 C07K016/12.

☐ 58. US20030118992A. A new ABC transporter polypeptide from Staphylococcus aureus is useful to diagnose, prevent or treat microbial infections and diseases. WARREN, R L. C07H021/04 C07K014/47 C12N005/06 C12P021/02 C12Q001/68 G01N033/48 G01N033/50 G06F019/00.

☐ 59. WO2003037310A. Composition useful for treating microbial infections, e.g. multidrug resistant microbial infections comprises opioid inhibitor of adenosine triphosphate-binding cassette drug transporter and anti-microbial agent. SCHOENHARD, G L. A61K031/00 A61K031/4709 A61K031/485 A61K031/496 A61K031/704 A61K031/7048 A61K031/7072 A61K031/7076 A61K038/13 A61K038/14.

☐ 60. WO 200234773A. Novel Streptococcus pneumoniae iron uptake ABC transporter peptide, useful in screening assay for identifying antimicrobial drug and in diagnostic assay for detecting streptococcal microorganism. BROWN, J S, et al. A61K039/09 C07K014/315 C07K016/12 C12N001/21 C12N015/63.

☐ 61. US20020081687A. Novel sitosterolemia susceptibility gene polypeptide and polynucleotide, useful for screening a compound that increases the level of expression or activity of SSG polypeptide for treating sterol-related disorder. SCHULTZ, J, et al. C07H021/04 C07K014/00 C12N005/06 C12N009/02 C12P021/02.

☐ 62. US 6251629B. New ABC transporter polypeptides and polynucleotides from Staphylococcus aureus, useful for screening antibiotics or as research reagents for diagnosing human diseases, e.g. emphysema, toxic shock syndrome or wound infection. WARREN, R L. C07H021/04 C12N015/31 C12N015/62 C12N015/63.

☐ 63. EP 1074623A. New Staphylococcus aureus ABC transporter polypeptides useful as diagnostic reagents, or for treating bacterial infections, e.g. otitis media, lung abscess, impetigo or wound infection. WARREN, R L. C07K014/31 C07K016/12 C12N015/31 C12Q001/68 G01N001/00.

☐ 64. US20020091092A. Use of propionyl L-carnitine or its salts for inducement of apoptosis used in the treatment of e.g. hypertension, pulmonary hypertension, the prevention of restenosis after angioplasty or coronary stenting, and for the treatment of tumors. CALVANI, M, et al. A61K031/00 A61K031/198 A61K031/205 A61K031/22 A61K031/223 A61K031/225 A61K031/47 A61K031/7024 A61K045/00 A61P009/00 A61P009/12 A61P035/00 A61P043/00.

☐ 65. WO 200027386A. Use of propionyl L- carnitine and its salts for inhibiting the proliferation of smooth muscular cells of the vascular wall used in e.g. the treatment of artherosclerosis, hypertension, pulmonary hypertension and preventing restenosis is new. CALVANI, M. A61K031/22.

☐ 66. US 6448224B. New nucleic acid and peptides, useful as antibiotic peptides. NOVAK, R, et al. A61K038/02 A61K038/16 A61K039/09 C07K007/08 C07K014/315 C07K016/12 C12N009/12 C12N009/14 C12N015/31 C12N015/54 C12N015/55 C12Q001/68.

☐ 67. WO 9950418A. Novel methods for the treatment and diagnosis of staphylococcal infections. BURNIE, J P. A61K031/7088 A61K038/00 A61K039/00 A61K039/085 A61K048/00 A61P031/04 C07K014/31 C07K016/00 C07K016/12 C12N015/09 C12N015/31 C12P021/02 C12Q001/04 C12Q001/68 G01N033/569 G01N033/68 C12N015/09 C12N015/09 C12N015/09 C12R001:44 C12R001:445 C12R001:45.

☐ 68. US 6300094B. New ABC transporter from Staphylococcus aureus - useful to treat infections. TRAINI, C M, et al. C07K014/31 C07K016/12 C12N001/15 C12N001/19 C12N001/21 C12N005/10 C12N015/09 C12N015/31 C12N015/63 C12N015/74 C12P021/02 C12Q001/68 G01N033/53 C12N015/09 C12R001:445.

☐ 69. EP 908516A. New polypeptides encoding members of the ABC transporter family from Staphylococcus aureus useful for diagnosing and treating diseases such as osteomyelitis and toxic shock syndrome. WARREN, R L. C07K014/31 C07K016/12 C12N015/31 C12Q001/68 G01N001/00.

☐ 70. WO 9801154A. Treating and diagnosing bacterial and fungal infection with ABC transporter protein - or neutralising or binding agents, and new Staphylococcal proteins, particularly for infections caused by drug resistant Staphylococci and Enterococci. BURNIE, J P, et al. A61K038/17 A61K039/00 A61K039/02 A61K039/085 A61K039/09 A61K039/395 A61K039/40 A61K045/00 A61P031/00 A61P031/04 C07H021/04 C07K014/31 C07K014/315 C07K014/435 C07K016/12 C12N001/18 C12N001/21 C12N005/06 C12N015/09 C12P021/02 G01N033/569.

☐ 71. EP 569541B. Treatment and repair of defects or lesions in cartilage - by a growth factor

containing matrix, esp. useful for treating osteoarthritis and traumatic cartilage damage. HUNZIKER, E B. A61F002/02 A61F002/30 A61K000/00 A61K009/00 A61K009/127 A61K009/70 A61K037/00 A61K037/02 A61K037/22 A61K037/24 A61K037/36 A61K037/48 A61K038/18 A61K038/27 A61K045/06 A61K047/00 A61K047/30 A61K047/42 C07K015/06 C09H003/02.

☐ 72. HU 45276T. Prepn. of oligomycin ABC complex by fermentation - with streptomyces diastatochromogenes var RO-31. ISTVAN, H, et al. C12P017/18.

☐ 73. US 3501568A. Antibiotic a3823 complex (factors a,b,c and d) anti- - bacterial antifungal anticoccidial antineoplastic. A61K021/00 C07G011/00.

Generate Collection Print

Terms	Documents
L4 and (strep$ or staph$ or (gram near3 positive))	73

Prev Page Next Page Go to Doc#

First Hit

Generate Collection | Print

L12: Entry 3 of 18 File: PGPB Jan 2, 2003

DOCUMENT-IDENTIFIER: US 20030004097 A1
TITLE: Methods and compositions for inducing autoimmunity in the treatment of cancers

Detail Description Paragraph:
[0178] Various methods have been used to determine the presence of PS on cell membranes. These include direct chemical modification with membrane impermeable reagents such as trinitrobenzenesulfonic acid and hydrolysis with specific phospholipases (Gordesky et al., 1975; Etemadi, 1980), direct labeling with PS binding proteins (Thiagarajan and Tait, 1990; Tait and Gibson, 1994; Vermes et al., 1995; Kuypers et al., 1996), and PS-dependent catalysis of coagulation (Rosing et al., 1980; Van Dieijen et al., 1981). Several laboratories used lipid antibodies to detect cell surface PS (Maneta-Peyret et al., 1993; Rote et al., 1993; Rote et al., 1995; Katsuragawa et al., 1995). However, many of these antibodies are not specific and cross-reactivity is common. This may be due to the weak antigenic presentation of the phosphorylated head groups that are critical to specificity or to the generation of antibodies to diacylglycerol, phosphodiester and/or fatty acid moieties that are common to all phospholipids. In an attempt to produce specific PS antibodies, the inventor immunized rabbits with PS covalently coupled to bovine serum albumin or KLH via its fatty acid side chain without modifying the crucial phosphoserine moiety.

L17: Entry 21 of 54 File: USPT Jan 2, 2001

DOCUMENT-IDENTIFIER: US 6168790 B1
**** See image for Certificate of Correction ****
TITLE: Use of antibodies to block the effects of gram-positive bacteria and mycobacteria

Detailed Description Text (163):
Heparinized whole mouse blood was distributed in a microtiter plate (200 .mu.l/well) and incubated in presence of LPS and polyclonal anti-murine CD14 IgG after 4 hours incubation at 37.degree. C., conditioned plasma were assayed for TNF bioactivity using the method of Espevik and Nissen-Meyer, supra. In experiments with THP-1 cells, cells were washed 2 times with serum-free RPMI containing 0.5 mg/ml human serum albumin, resuspended in serum free media, and distributed at the concentration of 5-7.times.10.sup.4 cells/well. Fetal bovine serum (Sigma) was added to obtain a final concentration of 5%. Various concentrations of LPS, cell wall preparations, LAM or soluble peptidoglycan were added to the cells with or without antibodies in duplicate, and incubated at 37.degree. C. for 7 hours. Cell free supernates were then sampled and frozen at -20.degree. C. IL-8 was measured with an ELISA as previously described by Standiford, et al. (J. Immunol., 145:1435-1439, 1990), with results as shown in FIG. 13.

First Hit Fwd Refs

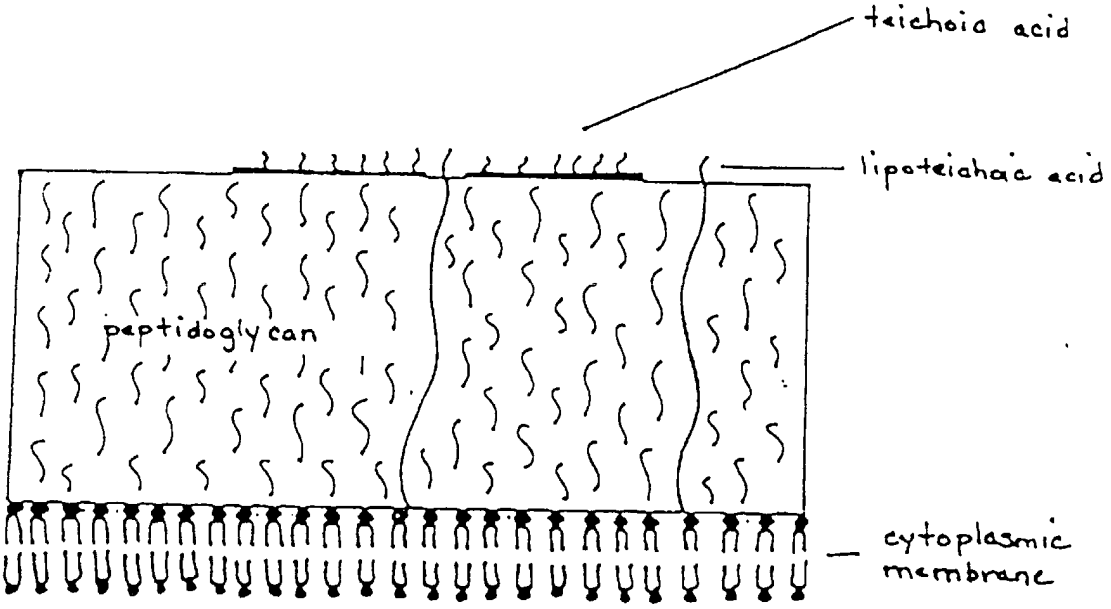
Generate Collection Print

L17: Entry 26 of 54 File: USPT Aug 15, 2000

DOCUMENT-IDENTIFIER: US 6103468 A
TITLE: Rapid two-stage polymerase chain reaction method for detection of lactic acid acid bacteria in beer

Brief Summary Text (24):
U.S. Pat. No. 5,139,933 discloses an assay method to quickly detect the presence of Listeria strains in samples, characterized by the use of antibodies to selectively capture the peptidoglycan and teichoic acid components of the listeriae bacterial cell wall.

FIGURE 1


teichoic acid
lipoteichoic acid
peptidoglycan
cytoplasmic membrane

YSTEM:OS - DIALOG OneSearch
File 155:MEDLINE(R) 1966-2004/May W5
(c) format only 2004 The Dialog Corp.
*File 155: Medline has been reloaded. Accession numbers
have changed. Please see HELP NEWS 154 for details.
File 348:EUROPEAN PATENTS 1978-2004/May W04
(c) 2004 European Patent Office
File 349:PCT FULLTEXT 1979-2002/UB=20040527,UT=20040520
(c) 2004 WIPO/Univentio
File 654:US Pat.Full. 1976-2004/Jun 01
(c) Format only 2004 The Dialog Corp.
*File 654: US published applications now online. See HELP NEWS 654
for details. Reassignments current through December 2, 2003.
File 5:Biosis Previews(R) 1969-2004/May W5
(c) 2004 BIOSIS
File 34:SciSearch(R) Cited Ref Sci 1990-2004/May W5
(c) 2004 Inst for Sci Info
File 35:Dissertation Abs Online 1861-2004/May
(c) 2004 ProQuest Info&Learning
File 73:EMBASE 1974-2004/May W5
(c) 2004 Elsevier Science B.V.
File 144:Pascal 1973-2004/May W4
(c) 2004 INIST/CNRS
File 357:Derwent Biotech Res. _1982-2004/Jun W1
(c) 2004 Thomson Derwent & ISI
File 440:Current Contents Search(R) 1990-2004/Jun 03
(c) 2004 Inst for Sci Info

Set	Items	Description

Cost is in DialUnits
?t s2/9/1 10

2/9/1 (Item 1 from file: 155)
DIALOG(R)File 155:MEDLINE(R)

09860124 PMID: 8212846
Assessment of non-protein impurities in potential vaccine proteins produced by Bacillus subtilis.
Himanen J P; Sarvas M; Helander I M
Department of Molecular Bacteriology, National Public Health Institute, Helsinki, Finland.
Vaccine (ENGLAND) 1993, 11 (9) p970-3, ISSN 0264-410X
Journal Code: 8406899
Document type: Journal Article
Languages: ENGLISH
Main Citation Owner: NLM
Record type: Completed
Subfile: INDEX MEDICUS
The levels of non-protein impurities at different stages of purification of model vaccine proteins produced by Bacillus subtilis were assessed with special emphasis on peptidoglycan-wall teichoic acid and **lipoteichoic** acid. Intracytoplasmically produced proteins were purified by disrupting the lysozyme protoplasts using osmotic shock, depositing the inclusion bodies by low-speed centrifugation, and washing them with detergent. By this procedure most of the cell envelope-derived impurities could be removed. The final product contained less than 1% (w/w) of neutral sugars, fatty acids, phosphate, **hexosamine** , diaminopimelic acid and glycerol. A secreted protein was purified from the culture supernatant by successive ion-exchange and adsorption chromatography. The cell envelope-derived impurities were efficiently removed by the cation-exchanger, and the final product contained only minute amounts of non-protein components. The amounts of non-protein components such as peptidoglycan and **lipoteichoic** acid in proteins produced in either mode were shown to be negligible in relation to their potentially harmful biological effects.
Tags: Comparative Study; Human; Support, Non-U.S. Gov't
Descriptors: *Bacillus subtilis--metabolism--ME; *Bacterial Outer Membrane Proteins--isolation and purification--IP; *Lipopolysaccharides

--analysis--AN; *Pertussis Toxin; *Recombinant Fusion Proteins--isolation and purification--IP; *Teichoic Acids--analysis--AN; *Vaccines, Synthetic --analysis--AN; *Virulence Factors, Bordetella--isolation and purification --IP; Carbohydrates--analysis--AN; Cell Fractionation--methods--MT; Chromatography--methods--MT; Detergents; Drug Contamination; Fatty Acids --analysis--AN; Phosphates--analysis--AN; Recombinant Fusion Proteins --immunology--IM; Vaccines, Synthetic--isolation and purification--IP
CAS Registry No.: 0 (Bacterial Outer Membrane Proteins); 0 (Carbohydrates); 0 (Detergents); 0 (Fatty Acids); 0 (Lipopolysaccharides); 0 (Phosphates); 0 (Recombinant Fusion Proteins) ; 0 (Teichoic Acids); 0 (Vaccines, Synthetic); 0 (Virulence Factors, Bordetella); 0 (pertussis toxin, S1 subunit); 0 (pertussis toxin, S4 subunit); 56411-57-5 (lipoteichoic acid)
Enzyme No.: EC 2.4.2.31 (Pertussis Toxin)
Record Date Created: 19931026
Record Date Completed: 19931026

2/9/10 (Item 1 from file: 35)
DIALOG(R)File 35:Dissertation Abs Online

01321483 ORDER NO: NOT AVAILABLE FROM UNIVERSITY MICROFILMS INT'L.
HETEROLOGOUS VACCINE PROTEINS PRODUCED IN BACILLUS SUBTILIS: CHEMICAL AND BIOLOGICAL ASSESSMENT OF HOST-DERIVED IMPURITIES (VACCINE)
Author: HIMANEN, JUHA-PEKKA
Degree: PH.D.
Year: 1992
Corporate Source/Institution: TURUN YLIOPISTO (FINLAND) (5760)
Source: VOLUME 54/04-C OF DISSERTATION ABSTRACTS INTERNATIONAL.
PAGE 1214. 59 PAGES
Descriptors: CHEMISTRY, BIOCHEMISTRY
Descriptor Codes: 0487
ISBN: 951-47-6548-6
Publisher: NATIONAL PUBLIC HEALTH INSTITUTE, MANNERHEIMINTIE 166, SF-00300 HELSINKI, FINLAND

In this study, a new strategy for producing safe and efficient vaccines was applied: the genes coding for proteins are cloned from the pathogen to Bacillus subtilis, which is non-toxic for humans and genetically well characterized. Three proteins were chosen for this study: soluble subunits S1 (BacS1) and S4 (BacS4) of pertussis toxin, and an outer membrane protein P1 of Neisseria meningitidis (BacP1). The proteins were produced either as secreted proteins or intracytoplasmically to cover different production systems.

Since there is only limited experience in using Bacillus subtilis for producing pharmaceuticals, no generally accepted methods for checking the bacillar non-protein impurities in the products are available. Therefore, the fractionation of the main non-protein cell envelop components of Bacillus subtilis, teichoic acids (TA's) and peptidoglycan (PG), during the purification of BacS1, BacS4, and BacP1 was investigated. The final preparations of BacS4 and BacP1 contained less than 1% (w/w) of all the tested non-protein components, i.e. neutral sugars, fatty acids, **hexosamines** , organic phosphate, DAP, and glycerol.

To estimate the biological significance of **lipoteichoic** acid (LTA) and peptidoglycanteichoic acid complex (PG-TA) in vaccine proteins, certain biological activities of purified and chemically characterized bacillar LTA and PG-TA were determined. LTA and PG-TA were found to be non-toxic for mice and guinea pigs in a short-term toxicity assay. PG-TA was weakly pyrogenic and mitogenic. Both LTA and PG-TA acted as immunologic adjuvants in mice. Both LTA and PG-TA, when injected in mice, also caused an increase in the number of granulocyte-monocyte colony-forming cells in the bone marrow probably via stimulation of production of granulocyte-monocyte colony-forming cells. With respect to vaccine production, the minute amounts of LTA and PG-TA in the purified heterologous proteins produced in Bacillus subtilis are not expected to cause harmful effects; instead, they may have beneficial effects on host defence. (Abstract shortened by UMI.)
?t s2/3,kwic/2 3 6 8 9

2/3,KWIC/2 (Item 1 from file: 348)
DIALOG(R)File 348:EUROPEAN PATENTS

00883189
IMMUNOMODULATORY COMPLEX AND USE THEREOF IN HELICOBACTER DISEASES
IMMUNMODULATORISCHER KOMPLEX UND DESSEN VERWENDUNG IN HELICOBACTER ERKRANKUNGEN
COMPLEXE IMMUNOMODULATEUR ET SON UTILISATION DANS LES AFFECTIONS PAR HELICOBACTER
PATENT ASSIGNEE:
TOROSSIAN, Fernand Narbey, (1858080), 10, rue Noel-Ballay,, F-31400 Toulouse, (FR), (Proprietor designated states: all)
INVENTOR:
TOROSSIAN, Fernand Narbey, 10, rue Noel-Ballay,, F-31400 Toulouse, (FR)
LEGAL REPRESENTATIVE:
Morelle, Guy Georges Alain (50595), Cabinet Morelle & Bardou, 9, Avenue de l'Europe BP 53, 31527 Ramonville Cedex, (FR)
PATENT (CC, No, Kind, Date): EP 969851 A1 000112 (Basic)
EP 969851 B1 040526
WO 1997030716 970828
APPLICATION (CC, No, Date): EP 97906249 970225; WO 97FR334 970225
PRIORITY (CC, No, Date): FR 962445 960226
DESIGNATED STATES: AT; BE; CH; DE; DK; ES; FI; FR; GB; GR; IE; IT; LI; LU; MC; NL; PT; SE
INTERNATIONAL PATENT CLASS: A61K-035/74; A61K-039/106; A61K-039/106; A61K-31:57; A61K-38:39; A61K-39:108; A61K-035/74; A61K-38:39
NOTE:
No A-document published by EPO
LANGUAGE (Publication,Procedural,Application): French; French; French
FULLTEXT AVAILABILITY:

Available Text	Language	Update	Word Count
CLAIMS B	(English)	200422	354
CLAIMS B	(German)	200422	348
CLAIMS B	(French)	200422	338
SPEC B	(French)	200422	3998
Total word count - document A			0
Total word count - document B			5038
Total word count - documents A + B			5038

...SPECIFICATION teichoic acids (Bacteriol. Reviews, 37, 21, 215-57).
G.A. MILLER (1976) - Effects of streptococcal **lipoteichoic** acid on host response in mice (Infect. and Immun., 1976, 13, (5), 1408-17).
A.J. WICKEN et coll. (1975) - **Lipoteichoic** acids: a new class of bacterial antigens (Science, 187, 1161-67). Differents dosages possibles A...
...Hexoses
T.A. SCOTT - Dosage colorimetr. a l'anthrone (Anal. Chem. (1953), 25, 1956-61).
Hexosamines
L.A. ELSON (Biochem. J (1953), 27, 1824-28).
Lipopolysaccharides
J. JANDA et E. WORK...

2/3,KWIC/3 (Item 2 from file: 348)
DIALOG(R)File 348:EUROPEAN PATENTS

00796491
ANTITUMOR PREPARATIONS CONTAINING A LIPOTEICHOIC ACID FROM STEPTOCOCCUS
ANTITUMORPRAPARATE, DIE EINE LIPOTEICHONSAURE AUS STREPTOCOCCUS ENTHALTEN
PREPARATIONS ANTITUMORALES CONTENANT UN ACIDE LIPOTEICHOIQUE TIRE DE STREPTOCOCCUS
PATENT ASSIGNEE:
Lunamed AG, (3874890), Kirschbaumweg 38, 4103 Bottmingen, (CH), (Proprietor designated states: all)
INVENTOR:
TRUOG, Peter, St. Johanns-Vorstadt 38, CH-4056 Basel, (CH)

ROTHLISBERGER, Peter, Breitensteinstrasse 100, CH-8037 Zurich, (CH)
LEGAL REPRESENTATIVE:
Arnold, Winfried et al (24682), Brugglistrasse 9, 4104 Oberwil, (CH)
PATENT (CC, No, Kind, Date): EP 807185 A1 971119 (Basic)
EP 807185 B1 020612
EP 807185 B9 021211
WO 9623896 960808
APPLICATION (CC, No, Date): EP 96902922 960124; WO 96EP309 960124
PRIORITY (CC, No, Date): EP 95101208 950130
DESIGNATED STATES: AT; BE; CH; DE; DK; ES; FR; GB; GR; IE; IT; LI; LU; MC; NL; PT; SE
INTERNATIONAL PATENT CLASS: C12P-019/44; C07H-015/04; C07H-015/06; A61K-031/702; A61K-031/70; C12N-001/20; C12R-001/46; A61K-038/46; C12N-001/20; C12R-1:46; C12P-019/44; C12R-1:46; A61K-031/702; A61K-38:00; A61K-031/702; A61K-38:46
NOTE:
No A-document published by EPO
LANGUAGE (Publication,Procedural,Application): English; English; English
FULLTEXT AVAILABILITY:

Available Text	Language	Update	Word Count
CLAIMS B	(English)	200250	352
CLAIMS B	(German)	200250	343
CLAIMS B	(French)	200250	434
SPEC B	(English)	200250	4453
Total word count - document A			0
Total word count - document B			5582
Total word count - documents A + B			5582

...SPECIFICATION new Streptococcus strain from which the new compound can be isolated.

Background of the Invention
Lipoteichoic acids (LTAs) are a group of amphipathic substances found in the cell wall of grain...

...and a hydrophobic glycolipid moiety. The hydrophilic backbone may be substituted with alanine, hexoses and **hexosamines** . The glycolipids described so far were mainly dihexosylglycerols and some trihexosylglycerols. **Lipoteichoic** acids show genus and species variation in the degree of polymerization of the hydrophilic chain...

...al., Science, 187, 1161 - 1167, (1975), and Microbiology, 360 - 365, (1977); Fischer W., Physiology of **lipoteichoic** acids in bacteria. Adv. Microb. Physiol., 29(233); 233-302 (1988), Fischer W., Mannsfeld T., Hagen G., On the basic structure of poly(glycerophosphate) **lipoteichoic** acids, Biochem. Cell Biol., 68 (1): 33-43, (1990).
LTAs have been reported as having...

2/3,KWIC/6 (Item 2 from file: 349)
DIALOG(R)File 349:PCT FULLTEXT

00341384 **Image available**
ANTITUMOR AND ANTICHOLESTEROL PREPARATIONS CONTAINING A LIPOTEICHOIC ACID FROM STEPTOCOCCUS
PREPARATIONS ANTITUMORALES ET ANTICHOLESTEROL CONTENANT UN ACIDE LIPOTEICHOIQUE TIRE DE STREPTOCOCCUS
Patent Applicant/Assignee:
TRUOG Peter,
RoTHLISBERGER Peter,
Inventor(s):
TRUOG Peter,
RoTHLISBERGER Peter,
Patent and Priority Information (Country, Number, Date):
Patent: WO 9623896 A1 19960808
Application: WO 96EP309 19960124 (PCT/WO EP9600309)
Priority Application: GB 95101208 19950130
Designated States: AL AM AU BB BG BR BY CA CN CZ EE FI GE HU IS JP KG KP KR

KZ LK LR MD MG MN MX NO NZ PL RU SG SK TJ TM TT UA US UZ VN KE LS MW SD
SZ UG AT BE CH DE DK ES FR GB GR IE IT LU MC NL PT SE BF BJ CF CG CI CM
GA GN ML MR NE SN TD TG
Publication Language: English
Fulltext Word Count: 6529
Fulltext Availability:
Detailed Description

Detailed Description
... new Streptococcus strain from which the new compound can be isolated

Background of the Invention
Lipoteichoic acids (LTAs) are a group of amphipathic sub stances found in the cell ...and a hydrophobic glycolipid moiety. The hydrophilic backbone may be substituted with alanine, hexoses and **hexosamines** . The glycolipids described so far were mainly dihexosylglycerols and some tri hexosylglycerols. **Lipoteichoic** acids show genus and species variation in the degree of polymerization of the hydrophilic chain...et al., Science, 187, - 1167, (1975), and Microbiology, 360 - 365, (1977); Fischer W., Physiology of **lipoteichoic** acids in bacteria. Adv. Microb. Physiol., 29(233): 233-302 (1988), Fischer W., Mannsfeld T., Hagen G., On the basic structure of poly (glycerophosphate) **lipoteichoic** acids, Biochem. Cell Biol., 68(1): 33-43, (1990)
LTAs have been reported as having...

2/3,KWIC/8 (Item 1 from file: 654)
DIALOG(R)File 654:US Pat.Full.

4488121 **IMAGE Available
Derwent Accession: 1996-371443
Utility
C/ **Antitumor and anticholesterol preparations containing a lipoteichoic acid from streptococcus**
; HAS LIPID ANCHOR, WHICH IS GALACTO-FURANOSYL-BETA-1-3-GLYCEROL WITH DIFFERENT FATTY ACIDS ESTERIFIED IN THE TWO ADJACENT HYDROXY GROUPS, AND HYDROPHILIC BACKBONE OF 10 GLYCEROPHOSPHATE UNITS ESTERIFIED WITH D-ALANINE TO 30%
Inventor: Truog, Peter, Basel, CH
Rothlisberger, Peter, Zurich, CH
Assignee: Lunamed AG (03), Bottmingen, CH
Lunamed AG CH (Code: 54449)
Examiner: Lee, Howard C. (Art Unit: 163)
Law Firm: Mathews, Collins, Shepherd & Gould, P.A.

	Publication Number	Kind	Date	Application Number	Filing Date
Main Patent	US 6214978	A	20010410	US 99232106	19990115
Division	Pending			US 97875018	19970623
Continuation	Pending			WO 96EP309	19960124
Priority				GB 95101208	19950130

Fulltext Word Count: 5326

Summary of the Invention:
... **Lipoteichoic** acids (LTAs) are a group of amphipathic substances found in the cell wall of gram...

...and a hydrophobic glycolipid moiety. The hydrophilic backbone may be substituted with alanine, hexoses and **hexosamines** . The glycolipids described so far were mainly dihexosylglycerols and some trihexosylglycerols. **Lipoteichoic** acids show genus and species

variation in the degree of polymerization of the hydrophilic chain...

...al., Science, 187, 1161-1167, (1975), and Microbiology, 360-365, (1977); Fischer W., Physiology of **lipoteichoic** acids in bacteria. Adv. Microb. Physiol., 29(233): 233-302 (1988), Fischer W., Mannsfeld T., Hagen G., On the basic structure of poly-(glycerophosphate) **lipoteichoic** acids, Biochem. Cell Biol., 68(1): 33-43, (1990...

2/3,KWIC/9 (Item 2 from file: 654)
DIALOG(R)File 654:US Pat.Full.

4377627
Derwent Accession: 1996-371443
Utility
C/ **Antitumor preparations**
; ISOLATED STREPTOCOCCUS SP PT STRAIN DSM 8747 WHICH PRODUCES A LIPOTEICHOIC ACID USED AS AN ANTICARCINOGENIC AGENT
Inventor: Truog, Peter, Basel, CH
Rothlisberger, Peter, Zurich, CH
Assignee: Lunamed AG (03), Bottmingen, CH
Lunamed AG CH (Code: 54449)
Examiner: Caputa, Anthony C. (Art Unit: 165)
Assistant Examiner: Weatherspoon, John K.
Law Firm: Mathews, Collins, Shepherd & Gould, P.A.

	Publication Number	Kind	Date	Application Number	Filing Date
Main Patent	US 6114161	A	20000905	US 97875018	19970623
PCT	WO 9623896		19960808	WO 96EP309	19960124
		371:	19971028		
		102e:	19971028		
Priority				EP 95101208	19950130

Fulltext Word Count: 5234

Summary of the Invention:
... **Lipoteichoic** acids (LTAs) are a group of amphipathic substances found in the cell wall of gram...

...and a hydrophobic glycolipid moiety. The hydrophilic backbone may be substituted with alanine, hexoses and **hexosamines** . The glycolipids de-scribed so far were mainly dihexosylglycerols and some trihexosylglycerols. **Lipoteichoic** acids show genus and species variation in the degree of polymerization of the hydrophilic chain...

...al., Science, 187, 1161-1167, (1975), and Microbiology, 360-365, (1977); Fischer W., Physiology of **lipoteichoic** acids in bacteria. Adv. Microb. Physiol., 29(233): 233-302 (1988), Fischer W., Mannsfeld T., Hagen G., On the basic structure of poly-(glycerophosphate) **lipoteichoic** acids, Biochem. Cell Biol., 68(1): 33-43, (1990
?logoff hold
03jun04 16:16:05 User228206 Session D2182.6
$0.06 0.020 DialUnits File155
$0.21 1 Type(s) in Format 9
$0.21 1 Types
$0.27 Estimated cost File155
$0.58 0.128 DialUnits File348
$3.40 2 Type(s) in Format 3
$3.40 2 Types
$3.98 Estimated cost File348
$0.42 0.088 DialUnits File349
$1.60 1 Type(s) in Format 3
$1.60 1 Types
$2.02 Estimated cost File349
$0.84 0.142 DialUnits File654

```
            $1.40  2 Type(s) in Format  3
          $1.40  2 Types
    $2.24  Estimated cost File654
          $0.04    0.007 DialUnits File5
    $0.04  Estimated cost File5
          $0.14    0.007 DialUnits File34
    $0.14  Estimated cost File34
          $0.08    0.020 DialUnits File35
            $2.30  1 Type(s) in Format  9
          $2.30  1 Types
    $2.38  Estimated cost File35
          $0.07    0.007 DialUnits File73
    $0.07  Estimated cost File73
          $0.02    0.007 DialUnits File144
    $0.02  Estimated cost File144
          $0.13    0.007 DialUnits File357
    $0.13  Estimated cost File357
          $0.14    0.007 DialUnits File440
    $0.14  Estimated cost File440
          OneSearch, 11 files,  0.439 DialUnits FileOS
    $0.24  TELNET
   $11.67  Estimated cost this search
   $11.67  Estimated total session cost   0.439 DialUnits


### Status: Signed Off. (1 minutes)
```

First Hit

Generate Collection

L1: Entry 1 of 1 File: DWPI Jan 1, 2003

DERWENT-ACC-NO: 1996-371443
DERWENT-WEEK: 200313

TITLE: New purified lipoteichoic acid cpd. LTA-T - isolated from new Streptococcus strain, used as antitumour and hypocholesterolaemic agent

INVENTOR: ROTHLISBERGER, P; TRUOG, P ; ROETHLISBERGER, P

PATENT-ASSIGNEE:

ASSIGNEE	CODE
TRUOG P	TRUOI
LUNAMED AG	LUNAN

PRIORITY-DATA: 1995EP-0101208 (January 30, 1995)

Search Selected | Search ALL | Clear

PATENT-FAMILY:

PUB-NO	PUB-DATE	LANGUAGE	PAGES	MAIN-IPC
ES 2178699 T3	January 1, 2003		000	C12P019/44
WO 9623896 A1	August 8, 1996	E	028	C12P019/44
AU 9647141 A	August 21, 1996		000	C12P019/44
EP 807185 A1	November 19, 1997	E	000	
JP 11500901 W	January 26, 1999		034	C12P019/44
US 6114161 A	September 5, 2000		000	C12N001/00
US 6214978 B1	April 10, 2001		000	C07H015/04
CN 1173206 A	February 11, 1998		000	C12P019/44
EP 807185 B1	June 12, 2002	E	000	C12P019/44
DE 69621776 E	July 18, 2002		000	C12P019/44
EP 807185 B9	December 11, 2002	E	000	C12P019/44

DESIGNATED-STATES: AL AM AU BB BG BR BY CA CN CZ EE FI GE HU IS JP KG KP KR KZ LK LR LR MD MG MN MX NO NZ PL RU SG SK TJ TM TT UA US UZ VN AT BE CH DE DK ES FR GB GR IE IT KE LS LU MC MW NL OA PT SD SE SZ UG AT BE CH DE DK ES FR GB GR IE IT LI LU MC NL PT SE AT BE CH DE DK ES FR GB GR IE IT LI LU MC NL PT SE AT BE CH DE DK ES FR GB GR IE IT LI LU MC NL PT SE

CITED-DOCUMENTS:7.Jnl.Ref; EP 135820 ; JP 61275217 ; US 3729461 ; WO 9420115

APPLICATION-DATA:

PUB-NO	APPL-DATE	APPL-NO	DESCRIPTOR
ES 2178699T3	January 24, 1996	1996EP-0902922	
ES 2178699T3		EP 807185	Based on
WO 9623896A1	January 24, 1996	1996WO-EP00309	
AU 9647141A	January 24, 1996	1996AU-0047141	
AU 9647141A		WO 9623896	Based on
EP 807185A1	January 24, 1996	1996EP-0902922	
EP 807185A1	January 24, 1996	1996WO-EP00309	
EP 807185A1		WO 9623896	Based on
JP 11500901W	January 24, 1996	1996JP-0523227	
JP 11500901W	January 24, 1996	1996WO-EP00309	
JP 11500901W		WO 9623896	Based on
US 6114161A	January 24, 1996	1996WO-EP00309	
US 6114161A	October 28, 1997	1997US-0875018	
US 6114161A		WO 9623896	Based on
US 6214978B1	January 24, 1996	1996WO-EP00309	Cont of
US 6214978B1	June 23, 1997	1997US-0875018	Div ex
US 6214978B1	January 15, 1999	1999US-0232106	
CN 1173206A	January 24, 1996	1996CN-0191668	
EP 807185B1	January 24, 1996	1996EP-0902922	
EP 807185B1	January 24, 1996	1996WO-EP00309	
EP 807185B1		WO 9623896	Based on
DE 69621776E	January 24, 1996	1996DE-0621776	
DE 69621776E	January 24, 1996	1996EP-0902922	
DE 69621776E	January 24, 1996	1996WO-EP00309	
DE 69621776E		EP 807185	Based on
DE 69621776E		WO 9623896	Based on
EP 807185B9	January 24, 1996	1996EP-0902922	
EP 807185B9	January 24, 1996	1996WO-EP00309	
EP 807185B9		WO 9623896	Based on

INT-CL (IPC): A61 K 31/70; A61 K 31/702; A61 K 31/71; A61 K 38/21; A61 K 38/46; C07 H 15/04; C07 H 15/06; C12 N 1/00; C12 N 1/04; C12 N 1/12; C12 N 1/20; C12 P 19/44; C12 R 1/46; C12 N 1/20; C12 R 1:46; C12 P 19/44; C12 R 1:46; A61 K 31/702; A61 K 38:00; A61 K 31/702; A61 K 38:46; C12 N 1/20; C12 R 1:46; C12 P 19/44; C12 R 1:46; A61 K 31/702; A61 K 38:00; A61 K 31/702; A61 K 38:46; C12 N 1/20; C12 R 1:46; C12 P 19/44; C12 R 1:46; A61 K 31/702; A61 K 38:00; A61 K 31/702; A61 K 38:46; C12 P 19/44; C12 R 1:01; C12 N 1/20; C12 R 1:01; C12 N 1/20; C12 R 1:46; C12 P 19/44; C12 R 1:46; A61 K 31/71; A61 K 38:00; A61 K 31/71; A61 K 38:46

ABSTRACTED-PUB-NO: EP 807185B
BASIC-ABSTRACT:

The following are new:

(1) a purified lipoteichoic acid (LTA), isolated from Streptococcus sp PT strain DSM 8747, designated 'LTA-T' and specifically of formula (I), and its salts;

(2) Streptococcus sp PT strain DSM 8747;

(3) a deacylated LTA cpd., designated 'LTA-T', of formula (I) in which both R1 and R2 are replaced by H, and its salt, and

(4) beta -galacto-furanosyl(1-3)glycerol-di-(R2) ester of formula (II) (opt. as a single cpd.) and its salt. R1 = H or D-alanyl with a ratio to P of 0.27-0.35; R2 = residues of satd. or unsatd. fatty acids with 12C, 14C, 16C or 18C; n = 9 (mean value).

USE - LTA-T has strong antitumour activity and is used for treating cancer (claimed), opt. in combination with a monokine and/or hyaluronidase. LTA-T is also used for lowering blood cholesterol levels (claimed). LTA-T and (II) are degradation degradation prods. of LTA-T and are useful as: (i) analytical tools for the identification and characterisation of LTA-T and (ii) starting materials for the prepn. of new LTA's with specific R2 gps. (e.g. by esterification of LTA-T with specific fatty acids) or with the 6-OH of the galacto-furanosyl moiety esterified by by a defined hydrophilic gp. LTA-T is administered orally or parenterally, esp. s.c., i.v. or i.p. at a concn. of 0.1-20 mu mol/ml. A typical antitumour dose is 0.001-20 (esp. 0.01-2) mg/kg, opt. in combination of 500-5000 (esp. ca 1000) U of hyaluronidase and/or 0.1-20 x 106 U of monokine.
ABSTRACTED-PUB-NO:

US 6114161A
EQUIVALENT-ABSTRACTS:

The following are new:

(1) a purified lipoteichoic acid (LTA), isolated from Streptococcus sp PT strain DSM DSM 8747, designated 'LTA-T' and specifically of formula (I), and its salts;

(2) Streptococcus sp PT strain DSM 8747;

(3) a deacylated LTA cpd., designated 'LTA-T', of formula (I) in which both R1 and R2 are replaced by H, and its salt, and

(4) beta -galacto-furanosyl(1-3)glycerol-di-(R2) ester of formula (II) (opt. as a single cpd.) and its salt. R1 = H or D-alanyl with a ratio to P of 0.27-0.35; R2 = residues of satd. or unsatd. fatty acids with 12C, 14C, 16C or 18C; n = 9 (mean value).

USE - LTA-T has strong antitumour activity and is used for treating cancer (claimed), opt. in combination with a monokine and/or hyaluronidase. LTA-T is also used for lowering blood cholesterol levels (claimed). LTA-T and (II) are degradation degradation prods. of LTA-T and are useful as: (i) analytical tools for the identification and characterisation of LTA-T and (ii) starting materials for the prepn. of new LTA's with specific R2 gps. (e.g. by esterification of LTA-T with specific fatty acids) or with the 6-OH of the galacto-furanosyl moiety esterified by by a defined hydrophilic gp. LTA-T is administered orally or parenterally, esp. s.c., i.v. or i.p. at a concn. of 0.1-20 mu mol/ml. A typical antitumour dose is 0.001-20 (esp. 0.01-2) mg/kg, opt. in combination of 500-5000 (esp. ca 1000) U of hyaluronidase and/or 0.1-20 x 106 U of monokine.

The following are new:

(1) a purified lipoteichoic acid (LTA), isolated from Streptococcus sp PT strain DSM DSM 8747, designated 'LTA-T' and specifically of formula (I), and its salts;

(2) Streptococcus sp PT strain DSM 8747;

(3) a deacylated LTA cpd., designated 'LTA-T', of formula (I) in which both R1 and

R2 are replaced by H, and its salt, and

(4) beta -galacto-furanosyl(1-3)glycerol-di-(R2) ester of formula (II) (opt. as a single cpd.) and its salt. R1 = H or D-alanyl with a ratio to P of 0.27-0.35; R2 = residues of satd. or unsatd. fatty acids with 12C, 14C, 16C or 18C; n = 9 (mean value).

USE - LTA-T has strong antitumour activity and is used for treating cancer (claimed), opt. in combination with a monokine and/or hyaluronidase. LTA-T is also used for lowering blood cholesterol levels (claimed). LTA-T and (II) are degradation degradation prods. of LTA-T and are useful as: (i) analytical tools for the identification and characterisation of LTA-T and (ii) starting materials for the prepn. of new LTA's with specific R2 gps. (e.g. by esterification of LTA-T with specific fatty acids) or with the 6-OH of the galacto-furanosyl moiety esterified by by a defined hydrophilic gp. LTA-T is administered orally or parenterally, esp. s.c., i.v. or i.p. at a concn. of 0.1-20 mu mol/ml. A typical antitumour dose is 0.001-20 (esp. 0.01-2) mg/kg, opt. in combination of 500-5000 (esp. ca 1000) U of hyaluronidase and/or 0.1-20 x 106 U of monokine.

US 6214978B

The following are new:

(1) a purified lipoteichoic acid (LTA), isolated from Streptococcus sp PT strain DSM DSM 8747, designated 'LTA-T' and specifically of formula (I), and its salts;

(2) Streptococcus sp PT strain DSM 8747;

(3) a deacylated LTA cpd., designated 'LTA-T', of formula (I) in which both R1 and R2 are replaced by H, and its salt, and

(4) beta -galacto-furanosyl(1-3)glycerol-di-(R2) ester of formula (II) (opt. as a single cpd.) and its salt. R1 = H or D-alanyl with a ratio to P of 0.27-0.35; R2 = residues of satd. or unsatd. fatty acids with 12C, 14C, 16C or 18C; n = 9 (mean value).

USE - LTA-T has strong antitumour activity and is used for treating cancer (claimed), opt. in combination with a monokine and/or hyaluronidase. LTA-T is also used for lowering blood cholesterol levels (claimed). LTA-T and (II) are degradation degradation prods. of LTA-T and are useful as: (i) analytical tools for the identification and characterisation of LTA-T and (ii) starting materials for the prepn. of new LTA's with specific R2 gps. (e.g. by esterification of LTA-T with specific fatty acids) or with the 6-OH of the galacto-furanosyl moiety esterified by by a defined hydrophilic gp. LTA-T is administered orally or parenterally, esp. s.c., i.v. or i.p. at a concn. of 0.1-20 mu mol/ml. A typical antitumour dose is 0.001-20 (esp. 0.01-2) mg/kg, opt. in combination of 500-5000 (esp. ca 1000) U of hyaluronidase and/or 0.1-20 x 106 U of monokine.

WO 9623896A

CHOSEN-DRAWING: Dwg.0/1

TITLE-TERMS: NEW PURIFICATION ACID COMPOUND ISOLATE NEW STREPTOCOCCUS STRAIN ANTITUMOUR HYPOCHOLESTEROLAEMIC AGENT

DERWENT-CLASS: B04 D16

CPI-CODES: B04-F10B4; B05-B01M; B07-A02A; B14-D02A2; B14-H01B; D05-C04;

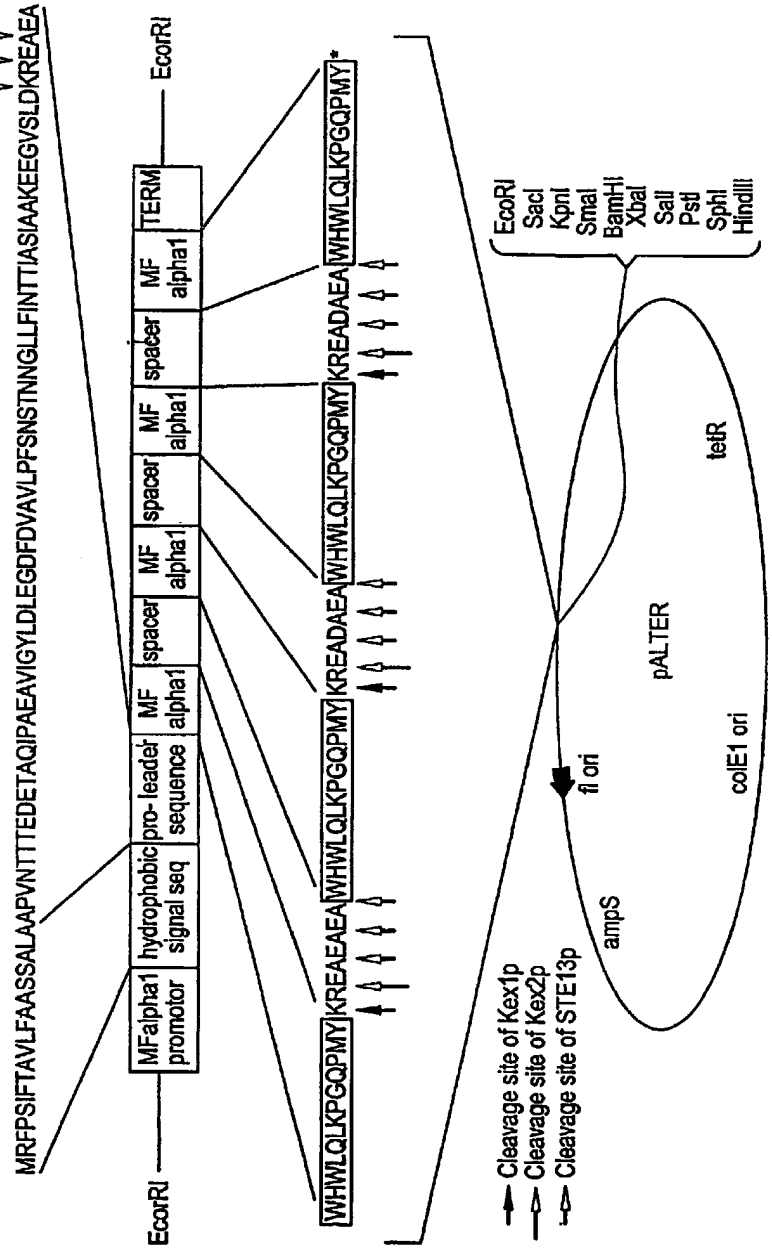
CHEMICAL-CODES:

Chemical Indexing M2 *01*
Fragmentation Code
B615 B702 B713 B720 B815 B833 F012 F013 F014 F015
F113 H100 H101 H181 H182 H4 H404 H405 H422 H482
H483 H484 H5 H521 H714 H721 H722 H8 J011 J012
J013 J014 J271 J272 J273 K0 L8 L815 L821 L833
M210 M211 M212 M213 M214 M215 M216 M220 M221 M222
M223 M224 M225 M226 M231 M232 M233 M262 M280 M281
M282 M312 M313 M321 M322 M323 M331 M332 M340 M342
M343 M349 M373 M381 M383 M391 M392 M393 M411 M510
M521 M530 M540 M710 M903 M904 N131 P831 Q233
Markush Compounds
199637-41801-N

Chemical Indexing M2 *02*
Fragmentation Code
F012 F013 F014 F015 F113 H4 H404 H405 H422 H482
H483 H484 H5 H521 H714 H721 H722 H8 J011 J012
J271 J272 K0 L8 L815 L821 L833 M210 M211 M212
M213 M214 M215 M216 M220 M221 M222 M223 M224 M225
M226 M231 M232 M233 M262 M280 M281 M282 M312 M313
M321 M332 M343 M373 M383 M391 M413 M510 M521 M530
M540 M710 M903 M904 N131 Q233
Markush Compounds
199637-41802-N

SECONDARY-ACC-NO:
CPI Secondary Accession Numbers: C1996-117941

FIG. 2


MRFPSIFTAVLFAASSALAAPVNTTTEDETAQIPAEAVIGYLDLEGDFDVAVLPFSNSTNNGLLFINTTIASIAAKEEGVSLDKREAEA
EcoRI
MFalpha1 promotor
hydrophobic signal seq
pro- leader sequence
MF alpha1
spacer
MF alpha1
spacer
MF alpha1
spacer
MF alpha1
TERM
EcoRI
WHWLQLKPGQPMY
KREAEAEA
WHWLQLKPGQPMY
KREADAEA
WHWLQLKPGQPMY
KREADAEA
WHWLQLKPGQPMY*
EcoRI
SacI
KpnI
SmaI
BamHI
XbaI
SalI
PstI
SphI
HindIII
Cleavage site of Kex1p
Cleavage site of Kex2p
Cleavage site of STE13p
ampS
f1 ori
pALTER
tetR
colE1 ori

TAC 3'. The 1.1 kb PCR product was gel purified with GeneClean II (Bio101), restricted with NcoI and BamHI and cloned into NcoI-BamHI cut and phosphatased Cadus 1122 to yield Cadus 1605. The sequence of Cadus 1605 was verified by restriction analysis and dideoxy-sequencing of double-stranded templates. Recombinant GPA_{Bam}-Gα_hybrids of Gαs, Gαi2, and Gα16 were generated. Construction of Cadus 1855 encoding recombinant GPA_{Bam}-Gα_16 serves as a master example: construction of the other hybrids followed an analogous cloning strategy. The parental plasmid Cadus 1617, encoding native Gα16, was restricted with NcoI and BamHI, treated with shrimp alkaline phosphatase as per the manufacturer's specifications and the linearized vector was purified by gel electrophoresis. Cadus 1605 was restricted with NcoI and BamHI and the 1.1 kb fragment encoding the amino terminal 60% of GPA1 with a novel BamHI site at the 3' end was cloned into the NcoI- and BamIII-restricted Cadus 1617. The resulting plasmid encoding the GPA_{Bam}-Gα_16 hybrid was verified by restriction analysis and assayed in tester strains for an ability to couple to yeast Gβγ and thereby suppress the gpa1 null phenotype. Two additional GPA_{Bam}-Gα_hybrids, GPA_{Bam}-Gαs_ and GPA_{Bam}-Gαi2, described in this application were prepared in an analogous manner using Cadus1606 as the parental plasmid for the construction of the GPA_{Bam}-Gα_i2 hybrid and Cadus 1181 as the parental plasmid for the construction of the GPA_{Bam}-Gα_s hybrid.

[0501] Coupling by chimeric Gα_proteins._The Gα chimeras described above were tested for the ability to couple a mammalian G protein-coupled receptor to the pheromone response pathway in yeast. The results of these experiments are outlined in Table 5. Results obtained using $GPA1_{41}$-Gαi2 to couple the human C5a receptor to the pheromone response pathway in autocrine strains of yeast are disclosed in Example 10 above.

TABLE 1

ABC TRANSPORTERS*

Species	System	Substrate
Bacteria		
Salmonella typhimurium	OppABCDF	Oligopeptides
Streptococcus pneumoniae	AmiABCDEF	Oligopeptides
Bacillus subtilis	Opp (SpoOK)	Oligopeptides
E. coli	Dpp	Dipeptides
Bacilus subtilis	DciA	Dipeptides
S. typhimurium	HisJQMP	Histidine
E. coli	HisJQMP	Histidine
E. coli	MalEFGK	Maltose
S. typhimurium	MalEFGK	Maltose
Enterobacter aerogenes	MalEFGK	Maltose
E. coli	UgpABCE	sn-Glycerol-3-phosphate
E. coli	AraFGH	Arabinose
E. coli	RbsACD	Ribose
E. coli	GlnHPQ	Glutamine
S. typhimurium	ProU (VWX)	Glycine-betaine
E. coli	ProU (VWX)	Glycine-betaine
E. coli	LivHMGF (JK)	Leucine-isoleucine-valine

TABLE 1-continued

ABC TRANSPORTERS*

Species	System	Substrate
E. coli	PstABC	Phosphate
Pseudomonas stutzeri	NosDYF	Copper
E. coli	ChlJD	Molybdenum
E. coli	CysPTWAM	Sulphate-Thiosulfate
E. coli	BtuCDE	Vitamin B12
E. coli	FhuBCD	Fe^{3+}-ferrichrome
E. coli	FecBCDE	Fe^{3+}-dicitrate
S. marcescens	SfuABC	Fe^{3+}
Mycoplasma	p37, 29, 69	?
E. coli	Phn/Psi	Alkyl-phosphonates (?)
Streptomyces peucetius	DrrAB	Daunomycin/Doxorubicin
Streptomyces fradiae	TlrC	Tylosin
Staphylococcus	MsrA	Erythromycin resistance
Agrobacterium tumefaciens	OccJQMP	Octopine
E. coli	HlyB	Haemolysin
Pasturella	LtkB	leukotoxin
E. coli	CvaB	Colicin V
Erwinia chrysanthemi	PrtD	Proteases
Bordetella pertussis	CyaB	Cyclolysin
Streptococcus pneumoniae	ComA	Competence factor
Rhizobium meliloti	NdvA	β-1,2-glucan
Agrobacterium tumefaciens	ChvA	β-1,2-glucan
Haemophilus influenzae	BexAB	Capsule polysaccharide
E. coli	KpsMT	Capsule polysaccharide
Niesseria	CrtCD	Capsule polysaccharide
E. coli	FtsE	Cell division
E. coli	UvrA	DNA repair
Rhizobium leguminosarum	NodI	Nodulation
Rhizobium meliloti	OFR1	?
Cyanobacteria		
Anabaena	HetA	Differentiation
Synchococcus	CysA	Sulphate
Yeast		
S. cerevisiae	STE6	a-mating peptide
S. cerevisiae	ADP1	?
S. cerevisiae	EF-3	Translation
Protozoa		
Plasmodium	pfMDR	Chloroquine
Lieshmania	ltpgpA	Methotrexate/heavy metals
Insect		
Drosophila	white-brown	Eye pigments
Drosophila	Mdr49	Hydrophobic drugs?
	Mdr65	?
Plants		
Liverwort chloroplast	MbpX	?

First Hit

L5: Entry 53 of 73 File: JPAB Sep 21, 1999

PUB-NO: JP411253178A
DOCUMENT-IDENTIFIER: JP 11253178 A
TITLE: ABC TRANSPORTER

PUBN-DATE: September 21, 1999

INVENTOR-INFORMATION:
NAME COUNTRY
WARREN, RICHARD L

ASSIGNEE-INFORMATION:
NAME COUNTRY
SMITHKLINE BEECHAM CORP

APPL-NO: JP10322783
APPL-DATE: October 7, 1998

INT-CL (IPC): C12 N 15/09; C07 K 14/31; C07 K 16/12; C12 P 21/02; C12 Q 1/68; G01 N 33/53; G01 N 33/577; A61 K 31/00; A61 K 38/00; A61 K 48/00

ABSTRACT:

PROBLEM TO BE SOLVED: To obtain a new isolated polypeptide comprising an ABC transporter containing an amino acid sequence having high identity of a specific amino acid sequence along the whole length, useful for searching an antibacterial compound, treating and diagnosing an infection with a bacterium of the genus Streptococcus, etc.

SOLUTION: This new ABC transporter comprises an isolated polypeptide containing an amino acid sequence having at least 70% identity of an amino acid sequence of the formula along the whole length of the formula and is useful for screening an antibacterial compound, treating and diagnosing an infection with Streptococcus aureus, etc. The isolated polypeptide is obtained by probing the library of a chromosome DNA clone of Streptococcus aureus WCUH29 with a partial sequence-derived radioactive labeled oligonucleotide, preferably an oligonucleotide of heptadecamer or longer than it, incorporating the obtained ABC transporter gene to an expression system and culturing a host cell.

Generate Collection Print

Search Results - Record(s) 51 through 73 of 73 returned.

☐ 51. 6077826. 08 Jun 98; 20 Jun 00. Synthetic macromolecular channel assembly for transport of chloride ions through epithelium useful in treating cystic fibrosis. Tomich; John M., et al. 514/12; 514/13 530/324 530/325 530/326. C07K014/00 A61K038/16.

☐ 52. 5773277. 18 May 95; 30 Jun 98. Crystalline chondroitinase isolated from Proteus vulgaris ATCC 6896. Hashimoto; Nobukazu, et al. 435/232; 435/183 435/188 435/201. C12N009/88 C12N009/44.

☐ 53. JP411253178A. 07 Oct 98. 21 Sep 99. ABC TRANSPORTER. WARREN, RICHARD L. C12N015/09; C07K014/31 C07K016/12 C12P021/02 C12Q001/68 G01N033/53 G01N033/577 A61K031/00 A61K038/00 A61K048/00.

☐ 54. WO003087147A1. 14 Apr 03. 23 Oct 03. STRETPTOCOCCAL GENES INVOLVED IN OSMOTIC AND OXIDATIVE STRESS AND IN VIRULENCE. BROWN, JEREMY STUART. C07K014/315; A61K035/74 C07K016/12.

☐ 55. WO009957281A2. 06 May 99. 11 Nov 99. PEPTIDE ANTIBIOTICS WHICH INHIBIT THE GROWTH OF PNEUMOCOCCI, ABC TRANSPORTER AND TWO-COMPONENT SIGNAL TRANSDUCTION SYSTEM PROTEINS FROM STREPTOCOCCUS PNEUMONIAE, AND METHODS USING THE SAME. NOVAK, RODGER, et al. C12N015/31; C12N015/54 C12N015/55 C12N009/12 C12N009/14 C07K007/08 C07K014/315 C07K016/12 C12Q001/68 A61K039/09.

☐ 56. WO009801154A2. 07 Jul 97. 15 Jan 98. TREATMENT AND DIAGNOSIS OF INFECTIONS OF GRAM POSITIVE COCCI. BURNIE, JAMES PETER, et al. A61K039/00; A61K039/02 A61K039/09 A61K039/40 G01N033/569 C07K014/31 C07K016/12 A61K039/085.

☐ 57. WO2003087147A. Peptides derived from phg ABC operon, useful for the manufacture of a medicament for treating or preventing a condition associated with infection by Streptococcus pneumoniae or other gram-positive bacteria. BROWN, J S. A61K035/74 C07K014/315 C07K016/12.

☐ 58. US20030118992A. A new ABC transporter polypeptide from Staphylococcus aureus is useful to diagnose, prevent or treat microbial infections and diseases. WARREN, R L. C07H021/04 C07K014/47 C12N005/06 C12P021/02 C12Q001/68 G01N033/48 G01N033/50 G06F019/00.

☐ 59. WO2003037310A. Composition useful for treating microbial infections, e.g. multidrug resistant microbial infections comprises opioid inhibitor of adenosine triphosphate-binding cassette drug transporter and anti-microbial agent. SCHOENHARD, G L. A61K031/00 A61K031/4709 A61K031/485 A61K031/496 A61K031/704 A61K031/7048 A61K031/7072 A61K031/7076 A61K038/13 A61K038/14.

☐ 60. WO 200234773A. Novel Streptococcus pneumoniae iron uptake ABC transporter peptide, useful in screening assay for identifying antimicrobial drug and in diagnostic assay for detecting streptococcal microorganism. BROWN, J S, et al. A61K039/09 C07K014/315 C07K016/12 C12N001/21 C12N015/63.

First Hit

End of Result Set

L5: Entry 73 of 73 File: DWPI

DERWENT-ACC-NO: 1970-18573R
DERWENT-WEEK: 197012

TITLE: Antibiotic a3823 complex (factors a,b,c and d) anti- - bacterial antifungal anticoccidial antineoplastic

Basic Abstract Text (1):
New antibiotic complex A3823 and its individual components, factors A, B, C and D. Factors A-D are caboxylic acids. A and B have the structure shown below. (A; R = Et, B; R = Me). Preparation of the complex is by aerobic fermentation of Streptomyces cinnamonensis (ATCC 15413).

```
File 135: New newsletters are now added. See Help News135 for the
complete list of newsletters.
  File 144:Pascal 1973-2004/May W5
         (c) 2004 INIST/CNRS
  File 149:TGG Health&Wellness DB(SM) 1976-2004/May W5
         (c) 2004 The Gale Group
  File 156:ToxFile 1965-2004/May W2
         (c) format only 2004 The Dialog Corporation
*File 156: ToxFile now reloaded with 2004 MeSH.
Enter Help News156 for more information.
  File 159:Cancerlit 1975-2002/Oct
         (c) format only 2002 Dialog Corporation
*File 159: Cancerlit ceases updating with immediate effect.
Please see HELP NEWS.
  File 162:Global Health 1983-2004/Apr
         (c) 2004 CAB International
  File 164:Allied & Complementary Medicine 1984-2004/May
         (c) 2004 BLHCIS
  File 172:EMBASE Alert 2004/May W5
         (c) 2004 Elsevier Science B.V.
  File 266:FEDRIP 2004/Apr
         Comp & dist by NTIS, Intl Copyright All Rights Res
  File 369:New Scientist 1994-2004/May W5
         (c) 2004 Reed Business Information Ltd.
  File 370:Science 1996-1999/Jul W3
         (c) 1999 AAAS
*File 370: This file is closed (no updates). Use File 47 for more current
information.
  File 399:CA SEARCH(R) 1967-2004/UD=14024
         (c) 2004 American Chemical Society
*File 399: Use is subject to the terms of your user/customer agreement.
Alert feature enhanced for multiple files, etc. See HELP ALERT.
  File 434:SciSearch(R) Cited Ref Sci 1974-1989/Dec
         (c) 1998 Inst for Sci Info
  File 444:New England Journal of Med. 1985-2004/Jun W1
         (c) 2004 Mass. Med. Soc.
  File 467:ExtraMED(tm) 2000/Dec
         (c) 2001 Informania Ltd.
*File 467: For information about updating status please see Help News467.

      Set  Items  Description
      ---  -----  -----------
Cost is in DialUnits
?ds

Set     Items   Description
S1          2   'ABC TRANSPORTER BCATRB' OR 'ABC-TYPE HEME TRANSPORTER, OR-
              F240-SUBUNIT, ORY'
S2       5446   R1-R2
S3       6030   R1:R6 OR R9
S4       6030   (S1 OR S2 OR S3)
S5     177536   'IMMUNOGLOBULIN' OR 'IMMUNOGLOBULIN A'
S6     109459   R1-R10
S7     437613   E1-E12
S8     609571   R1-R3
S9     415866   E1-E12
S10    758023   ANTIBOD? OR IGG OR IGM OR IMMUNOGLOBU? OR MONOCLONAL? OR M-
              AB OR MOAB OR HUMANIZ?
S11       637   S4 AND (S5 OR S6 OR S7 OR S8 OR S9 OR S10)
S12       148   S11 AND (TREAT? OR PREVENT? OR THERAP?)
S13         6   S12 AND (STAPH? OR STREP? OR (GRAM (2N) POSITIVE))
S14     14890   ATP(N)BINDING (N) CASSETTE? (N) TRANPORT? OR (ABC (2N) TRA-
              NSPORT?)
S15   4397397   ANTIBOD? OR IMMUNOGLOB? OR MONOCLONAL? OR SCFV? OR AB OR M-
              OAB OR MAB OR HUMANIZ? OR CHIMERIC? OR CHIMAER? OR (SINGLE (2-
              N) CHAIN)
S16  20455382   TREAT? OR PREVENT? OR IMMUNOPASSIV? OR IMMUNOTHER? OR THER-
              AP? OR IVIG OR IGIV OR IGGIV OR IVIGG
S17         2   S1 AND S2 AND S3
```

```
S18       4252   S4/1998:2004
S19       5981   S4 NOT S5
?t s17/6/all
```

```
Cost is in DialUnits
?ds

Set     Items   Description
S1          2   'ABC TRANSPORTER BCATRB' OR 'ABC-TYPE HEME TRANSPORTER, OR-
              F240-SUBUNIT, ORY'
S2       5446   R1-R2
S3       6030   R1:R6 OR R9
S4       6030   (S1 OR S2 OR S3)
S5     177536   'IMMUNOGLOBULIN' OR 'IMMUNOGLOBULIN A'
S6     109459   R1-R10
S7     437613   E1-E12
S8     609571   R1-R3
S9     415866   E1-E12
S10    758023   ANTIBOD? OR IGG OR IGM OR IMMUNOGLOBU? OR MONOCLONAL? OR M-
              AB OR MOAB OR HUMANIZ?
S11       637   S4 AND (S5 OR S6 OR S7 OR S8 OR S9 OR S10)
S12       148   S11 AND (TREAT? OR PREVENT? OR THERAP?)
S13         6   S12 AND (STAPH? OR STREP? OR (GRAM (2N) POSITIVE))
S14     14890   ATP(N)BINDING (N) CASSETTE? (N) TRANPORT? OR (ABC (2N) TRA-
              NSPORT?)
S15   4397397   ANTIBOD? OR IMMUNOGLOB? OR MONOCLONAL? OR SCFV? OR AB OR M-
              OAB OR MAB OR HUMANIZ? OR CHIMERIC? OR CHIMAER? OR (SINGLE (2-
              N) CHAIN)
S16  20455382   TREAT? OR PREVENT? OR IMMUNOPASSIV? OR IMMUNOTHER? OR THER-
              AP? OR IVIG OR IGIV OR IGGIV OR IVIGG
S17         2   S1 AND S2 AND S3
S18      4252   S4/1998:2004
S19      5981   S4 NOT S5
?s s14 and s15 and s16
Processing
Processed  10 of  26 files ...
Completed processing all files
         14890  S14
       4397397  S15
      20455382  S16
    S20     318  S14 AND S15 AND S16
?ds

Set     Items   Description
S1          2   'ABC TRANSPORTER BCATRB' OR 'ABC-TYPE HEME TRANSPORTER, OR-
              F240-SUBUNIT, ORY'
S2       5446   R1-R2
S3       6030   R1:R6 OR R9
S4       6030   (S1 OR S2 OR S3)
S5     177536   'IMMUNOGLOBULIN' OR 'IMMUNOGLOBULIN A'
S6     109459   R1-R10
S7     437613   E1-E12
S8     609571   R1-R3
S9     415866   E1-E12
S10    758023   ANTIBOD? OR IGG OR IGM OR IMMUNOGLOBU? OR MONOCLONAL? OR M-
              AB OR MOAB OR HUMANIZ?
S11       637   S4 AND (S5 OR S6 OR S7 OR S8 OR S9 OR S10)
S12       148   S11 AND (TREAT? OR PREVENT? OR THERAP?)
S13         6   S12 AND (STAPH? OR STREP? OR (GRAM (2N) POSITIVE))
S14     14890   ATP(N)BINDING (N) CASSETTE? (N) TRANPORT? OR (ABC (2N) TRA-
              NSPORT?)
S15   4397397   ANTIBOD? OR IMMUNOGLOB? OR MONOCLONAL? OR SCFV? OR AB OR M-
              OAB OR MAB OR HUMANIZ? OR CHIMERIC? OR CHIMAER? OR (SINGLE (2-
              N) CHAIN)
S16  20455382   TREAT? OR PREVENT? OR IMMUNOPASSIV? OR IMMUNOTHER? OR THER-
              AP? OR IVIG OR IGIV OR IGGIV OR IVIGG
S17         2   S1 AND S2 AND S3
S18      4252   S4/1998:2004
S19      5981   S4 NOT S5
S20       318   S14 AND S15 AND S16
?s s20/1998:2004
Processing
Processed  10 of  26 files ...
```

```
>>>One or more prefixes are unsupported
>>>  or undefined in one or more files.
>>>Year ranges not supported in one or more files
Completed processing all files
             314  S20
        32054088  PY=1998 : PY=2004
     S21     280  S20/1998:2004
?s s20 not s21
             318  S20
             280  S21
     S22      38  S20 NOT S21
?t s22/6/all
```

MOLECULAR CHARACTERIZATION OF STREPTOCOCCUS UBERIS CAMP FACTOR, LACTOFERRIN BINDING PROTEIN AND THEIR UPSTREAM GENES

Author: JIANG, MIN
Degree: PH.D.
Year: 1996
Corporate Source/Institution: THE UNIVERSITY OF SASKATCHEWAN (CANADA) (0780)
Adviser: L. A. BABIUK
Source: VOLUME 58/12-B OF DISSERTATION ABSTRACTS INTERNATIONAL. PAGE 6431. 234 PAGES
Descriptors: BIOLOGY, VETERINARY SCIENCE ; BIOLOGY, MOLECULAR ; BIOLOGY, MICROBIOLOGY
Descriptor Codes: 0778; 0307; 0410
ISBN: 0-612-24025-8

The gene coding for the CAMP factor from a strain of Streptococcus uberis (S. uberis) was cloned in E. coli. Chromosomal DNA from S. uberis was used to construct a gene library in plasmid pTZ18R and six CAMP-reaction positive clones were obtained from a total of 10,000 transformants. One clone, pJLD21, was subcloned and the CAMP factor gene (cfu) was localized to a 3.2 kb BamHI fragment. The nucleotide sequence of cfu was determined and the deduced amino acid sequence shown to be homologous to the corresponding Streptococcus agalactiae (S. agalactiae) protein. Immunoblot analysis revealed that the recombinant strain containing pJLD21 expressed a protein with a molecular weight of 28,000. **Antibodies** raised against purified S. uberis CAMP factor cross-reacted with S. agalactiae protein B. Southern blot analysis demonstrated that the six CAMP-reaction positive E. coli clones contained the same CAMP factor gene, and this gene existed in three out of eight S. uberis strains.

An ORF encoding a 277-residue protein was identified upstream of the CAMP factor gene. Sequence analysis indicated that the gene product is potentially a polar amino acid and opine binding protein of an **ABC** -type **transport** system.

The interaction between S. uberis and bovine lactoferrin (bLf) has been characterized. Apo-bLf could inhibit $\sp{125}$I-bLf binding as effectively as iron-saturated bLf. Bovine transferrin, human lactoferrin and human transferrin did not interfere with bLf binding. The Scatchard plot was linear and approximately 7800 binding sites were expressed by each bacterial cell, with an affinity of $1.0\times 10\sp{-7}$ M. Heat- or protease **treatment** of bacterial cells reduced bLf binding to a great degree. Two components with estimated molecular weights of 165,000 and 76,000 were originally identified from the cell wall as the functionally active bLf binding proteins.

The gene coding for the bLf binding protein (Ibp) of S. uberis has been cloned and sequenced. A single ORF encoding 561 amino acid residues resulted in the presence of two proteins in the recombinant E. coli cell. These proteins were able to bind bovine lactoferrin and had molecular weights of 76,000 and 165,000, similar to those detected in S. uberis. A putative signal peptide was found at the N terminus of the deduced amino acid sequence and the C terminus had the features of the membrane anchor motif found in other surface proteins from Gram positive bacteria. Deletion analysis located the bLf binding domain to a 200 amino acid region at the N terminus of this protein.

The vaccine potential of recombinant CAMP factor and lactoferrin binding protein has been evaluated. (Abstract shortened by UMI.)

Higgins CF. 1992. **ABC transporters** : from microorganisms to man. Annu. Rev. Cell Biol. 8:67-113

03293644 H.W. WILSON RECORD NUMBER: BGSA96043644 (THIS IS THE FULLTEXT)

The biochemistry and genetics of capsular polysaccharide production in bacteria.

Roberts, Ian S

Annual Review of Microbiology v. 50 (1996) p. 285-315
SPECIAL FEATURES: bibl il ISSN: 0066-4227
LANGUAGE: English
COUNTRY OF PUBLICATION: United States
RECORD TYPE: Abstract; Fulltext RECORD STATUS: Corrected or revised record
WORD COUNT: 14812

ABSTRACT: Bacterial polysaccharides are usually associated with the outer surface of the bacterium. They can form an amorphous layer of extracellular polysaccharide (EPS) surrounding the cell that may be further organized into a distinct structure termed a capsule. Additional polysaccharide molecules such as lipopolysaccharide (LPS) or lipooligosaccharide (LOS) may also decorate the cell surface. Polysaccharide capsules may mediate a number of biological processes, including invasive infections of human beings. Discussed here are the genetics and biochemistry of selected bacterial capsular polysaccharides and the basis of capsule diversity but not the genetics and biochemistry of LPS biosynthesis (for reviews see 100, 140). Reprinted by permission of the publisher.

TEXT:

INTRODUCTION

Polysaccharide capsules are ubiquitous structures found on the cell surface of a broad range of bacterial species. The polysaccharide capsule often constitutes the outermost layer of the cell; as such, it may mediate direct interactions between the bacterium and its immediate environment and has been implicated as an important factor in the virulence of many animal and plant pathogens (23, 82, 103). Capsular polysaccharides are linked to the cell surface of the bacterium via covalent attachments to either phospholipid or lipid-A molecules (140). In contrast, extracellular polysaccharide (EPS) molecules appear to be released onto the cell surface with no visible means of attachment and are often sloughed off to form slime. The release of polysaccharide from the cell surface must be used with caution as a criterion for differentiating between capsules and EPS. Capsular polysaccharides may themselves be released into the growth medium as a consequence of the stability of the phosphodiester linkage between the polysaccharide and the phospholipid membrane anchor. In addition, certain EPS molecules appear to remain tightly associated with the cell surface in the absence of detectable membrane anchoring (127).

Capsular polysaccharides are highly hydrated molecules that are over 95[percent] water (22). They are composed of repeating single units (monosaccharides) joined by glycosidic linkages. They can be homo- or heteropolymers and may be substituted by both organic and inorganic molecules. Any two monosaccharides may be joined in a number of configurations as a consequence of the multiple hydroxyl groups within each monosaccharide unit that may be involved in the formation of a glycosidic bond. As a result of this, capsular polysaccharides are an incredibly diverse range of molecules that may differ not only by monosaccharide units but also in how these units are joined together. The introduction of branches into the polysaccharide chain and substitution of both organic and inorganic molecules yield additional structural complexity.

In the case of human pathogens, a large number of different capsule serotypes have been identified. Over 80 different capsular polysaccharides or K antigens have been described for Escherichia coli, of which a small fraction are associated with invasive infections (87). The expression of particular K antigens can be associated with specific infections. For instance, E. coli that expresses the K1 antigen, a homopolymer of a2,8-linked N-acetylneuraminic acid (NeuNAc), is the major cause of neonatal meningitis (101). Certain E. coli K antigens have identical polysaccharide chains and differ only in modification of the

polysaccharide. The K13, K20, and K23 antigens are all polymers of ribose (rib) and 2-keto-3-deoxymanno-octonic acid (KDO) that have the structure = 3)-b-D-Rib-(1-7)-b-KDO-(2 = (131, 129)). However, in the case of the K13 and K20 antigens, the molecule is O-acetylated at the KDO and rib units respectively (129, 131). Chemically identical capsular polysaccharides may also be synthesized by different bacterial species. The Neisseria meningitidis group B capsular polysaccharide is identical to the K1 polymer of E. coli (48), and the E. coli K18, K22, and K100 antigens have the same constituents and a structure similar to serotype b capsule of Haemophilus influenzae (59). The apparent conservation of particular capsular polysaccharide structures between taxonomically diverse bacterial general raises interesting questions concerning the evolution of capsule diversity and the acquisition and transmission of capsule biosynthesis genes.

FUNCTIONS OF BACTERIAL CAPSULES

A number of possible functions have been suggested for polysaccharide capsules (Table 1).

PREVENTION OF DESICCATION

The formation by capsules of a hydrated gel around the surface of the bacteria may protect the bacteria from the harmful effects of desiccation (102). This may be particularly relevant in aiding the transmission of encapsulated pathogens from one host to the next. Mucoid isolates of E. coli, Acinetobacter calcoaceticus, and Erwinia stewartii are more resistant to drying than are isogenic nonmucoid strains (85). In the case of E. coli, desiccation increases expression of the genes encoding enzymes for the biosynthesis of colanic add (cell-surface slime) (85). The mechanism by which bacteria might regulate capsule expression in response to desiccation is unclear. One possible mechanism is that desiccation changes the external osmolarity, which triggers increased capsule biosynthesis. Indeed, both alginate production by Pseudomonas aeruginosa (9) and expression of the Vi antigen in Salmonella typhi increase in response to high external osmolarity (98).

ADHERENCE

Capsular polysaccharides may promote the adherence of bacteria to both surfaces and to each other, and they thereby facilitate the formation of a biofilm and the colonization of various ecological niches (21). During colonization of oral surfaces and the temporal development of bacterial plaque, specific colonizing bacteria may provide bridges for the subsequent attachment of other bacterial species (66). This intergeneric interaction that establishes microbial consortia within the biofilm is mediated in part through lectin-ligand interactions that involve cell-surface polysaccharide molecules (64). The formation of biofilms may offer the individual bacteria protection from phagocytic protozoa and infection by bacteriophages as well as nutritional advantages.

The ability of bacteria to attach to surfaces and establish a biofilm can have far-reaching consequences. The colonization of indwelling catheters in hospitalized patients can lead to serious nosocomial infections. The overexpression of alginate by Pseudomonas species within the lungs of cystic fibrosis patients forms an alginate-rich biofilm that may present a permeability barrier to antibiotics (47). The fouling of pipes in industrial processes due to biofilm formation can lead to substantial economic losses and delays in production while biofilms are removed (21). In contrast to these adherent properties, capsular polysaccharides may have lubricant properties that facilitate the swarming of Proteus mirabilis over solid substrata by reducing friction (50).

RESISTANCE TO NONSPECIFIC HOST IMMUNITY

During invasive bacterial infections, interactions between the capsular polysaccharide and the host's immune system can decide the outcome of the infection (105). In the absense of specific **antibody** , the presence of a capsule is thought to confer resistance to nonspecific host defense mechanisms. These responses include activation of the complement cascade via the alternative pathway and of the C3b-mediated opsonophagocytosis by polymorphonuclear leukocytes. Both of these responses provide protection

in the preimmune host when specific **antibodies** are absent. The alternative pathway is initiated by the nonspecific binding of the serum protein C3b to the bacterial cell surface. The bound C3b is then activated by interaction with factor B and forms the C3 convertase C3bBb. This leads to the binding of more C3 and the formation of the membrane attack complex on the outer membrane of the bacteria, which leads to lysis and death (38). The capsule may resist complement-mediated killing by providing a permeability barrier to complement components, thereby masking underlying cell surface structures that would otherwise be potent activators of the alternative pathway (56).

Capsular polysaccharides containing NeuNAc are poor activators of the alternative pathway (32, 80, 122). Poor activation may be achieved by NeuNAc containing polysaccharides directly binding factor H (80). The bound factor H acts as a cofactor to promote the binding of factor I with C3b to form iC3b, which breaks the amplification loop of the complement cascade and thereby **prevents** the formation of the membrane attack complex (38). The capsule usually acts in concert with other cell-surface structures such as O antigens to confer resistance to complement-mediated killing (65). As a result, a particular combination of cell-surface structures is responsible for conferring a high degree of resistance to complement-mediated killing (23, 65).

Capsular polysaccharides may confer resistance to complement-mediated opsonophagocytosis. Steric effects, in which the capsule masks the underlying C3b deposited on cell surface structures from C3b receptors on the phagocyte cell surface, may be responsible. The net negative charge conveyed on the cell surface by the polysaccharide capsule may also serve to confer resistance (16, 55, 82). The more highly charged the capsular polysaccharide is, the greater the degree of resistance to phagocytosis (82). In addition to these direct interactions between the bacterial capsule and components of the host's nonspecific immune response, certain capsular polysaccharides may modulate the ability of the host to mediate an immune response by effecting the release of cytokine molecules, thereby disrupting the coordination of the host's cell-mediated immune response (23).

RESISTANCE TO SPECIFIC HOST IMMUNITY

Although most capsular polysaccharides can elicit an immune response, a small set of capsular polysaccharides are poorly immunogenic. These include polysaccharides containing NeuNAc, such as E. coli K1 or N. meningitidis serogroup B (10), and the E. coli K5 antigen which is similar to desulfoheparin (130). As a consequence of structural similarities between these capsular polysaccharides and polysaccharides encountered on host tissue (37, 76), these capsules are poorly immunogenic, and infected individuals mount a poor **antibody** response to such capsules (105, 142). Therefore, the expression of these capsules confers some measure of resistance to the host's specific immune response.

BACTERIA-PLANT INTERACTIONS

Capsular polysaccharides play important roles in the mediation of microbeplant interactions. Many phytopathogenic bacteria elaborate capsular polysaccharides, the expression of which is essential for virulence. Unlike their role in human and animal pathogens, the role of capsular polysaccharides in the plant disease process has not been fully elucidated. In the case of certain vascular pathogens such as Pseudomonas solanacearum, the capsular polysaccharide is not required for invasion and growth of the bacteria in planta but is necessary for plant death, which is probably caused by occlusion of the xylem vessels (24). With other phytopathogens such as Erwinia amylovora, the expression of a polysaccharide capsule is essential for growth in planta and may act to mask cell-surface molecules that might otherwise elicit a defense response in the plant (18).

The expression of capsular polysaccharides is vital in the establishment of symbiotic relationships between bacteria and plants. Rhizobium meliloti, which infects leguminous plants to form nitrogen-fixing root nodules, produces succinoglycan, a cell-surface polymer made up of octasaccharide units--each composed of one galactose and seven glucose

residues with acetyl, succinyl, and pyruvyl substitutions (45). Mutants of R. meliloti that are unable to make succinoglycan can induce nodule formation on alfalfa but do not penetrate or colonize the nodule, which indicates a role for succinoglycan in nodule invasion and development (84).

GENETICS OF CAPSULE GENE CLUSTERS IN GRAM-NEGATIVE BACTERIA

Capsule gene clusters have been cloned from a number of gram-negative bacteria, including E. coli (104, 118), H. influenzae (70), N. meningitidis (42), S. typhi (53), P. solanacearum (57), Klebsiella pneumoniae (3), Erwinia stewartii (29), and E. amylovora (18). In all these cases the capsule genes are clustered at a single chromosomal locus, which allows the coordinate regulation of a large number of genes that may be involved in the biosynthesis and export of capsular polysaccharides. The capsule gene clusters of E. coli, the most studied forms to date, may be regarded as a paradigm for capsule gene clusters in gram-negative bacteria.

CAPSULE GENE CLUSTERS IN E. COLI

Escherichia coli can produce over 80 chemically distinct capsular polysaccharides (K antigens) (87). The K antigens of E. coli were divided into groups I and II on the basis of their biological and chemical properties (59). The two-group classification now appears to be inadequate for a number of reasons. First, group I K antigens include two separate capsule types (see below). Second, at the genetic level, at least two different groups of E. coli capsule gene clusters have recently been identified at the serA locus on the E. coli chromosome (96).

GROUP I K ANTIGENS

Group I K antigens have higher molecular weights than those of group II and lower charge density, are expressed at all growth temperatures (59), and produce thicker capsules (7). They contain hexuronic acid as acidic components, may contain amino sugars, and are coexpressed, usually with O8 or O9 and rarely O20 antigens of E. coli (62). Group I K antigens have been subdivided into groups Ia and Ib based on the presence of amino sugars (62). Group Ia K antigens do not contain amino sugars, and they resemble the capsular polysaccharides of Klebsiella species. Strains expressing group Ia K antigens are unable to express cell surface-colanic acid or M antigen, which suggests that these genes may be allelic (63). In contrast, group Ib K antigens which contain amino sugars and have no obvious counterpart in other bacteria are able to express colanic acid (63). Group I K antigens are linked to the cell surface by lipid A core in a manner analogous to that of lipopolysaccharide (LPS) (60). However, careful analysis of both group Ia (K30) and Ib (K40) polysaccharides indicates that the situation is more complex. Analysis of the size of the polysaccharide attached to lipid A-core reveals differences between group Ia and Ib K antigens. In the case of the K30 antigen, the polysaccharide linked to lipid A-core consists primarily of one repeat unit of the K30 polysaccharide with a high molecular weight polysaccharide unlinked to lipid A-core (77). This, together with the observation that in rfa mutants, which are unable to make lipid A-core, a K30 capsule can still be visualized by electron microscopy (77), indicates that either the polysaccharide is held on the cell surface independent of any covalent attachment or is anchored by some other molecule apart from either lipid A-core or phospholipid. The presence of both lipid A-core substituted and unsubstituted forms of the K30 antigen on the cell surface may indicate that these two different species are exported from the cell by different pathways.

In the case of the E. coli K40 antigen (group Ib), the majority of the capsular material consists of high molecular weight polysaccharide chains that are linked to lipid A core; only a small fraction are not substituted with lipid A-core (27). This is much more typical of an LPS molecule. Further evidence lends support to the notion that group Ib K antigens should be considered as O antigens: first, the effect of the rol gene product on the expression of group Ib K antigens. The rol gene controls the length of polymerization of heteropolymeric LPS O antigens in a range of Enterobacteriacae (5, 6, 81). In the case of group Ib K antigens, the rol gene is present on the chromosome, and multiple copies of the rol gene reduce the chain length of group Ib K antigens in a way analogous to the

effects on O antigen chain length (27). Group Ia strains lack a detectable rol gene. Second, in different E. coli strains, chemically identical polysaccharides are serologically classified as either an O or a group Ib K antigen (59). On the basis of this evidence, the classification of group I K antigens of E. coli may be too rigid.

The genes for the production of group Ia K antigens are located proximal to the his and near the rfb gene clusters on the E. coli chromosome (72, 138). The genes for group Ib K antigens have been assumed to be located at the same region of the chromosome, but this has not yet been demonstrated. A second locus near the trp gene cluster has also been implicated in the expression of the K27 antigen (112), although the role of this trp-linked marker in the expression of other group Ia K antigens awaits elucidation. Part of the cps gene cluster for the production of colanic acid in E. coli has been cloned and analyzed (2). To date, this represents the only group Ia capsule gene cluster that has been cloned from E. coli. Analysis of the sequence identified six open reading frames (ORFs). The two genes cpsB and cpsG encode the enzymes mannose-1-phosphate guanyl-transferase and phosphomannomutase respectively, and together with the other ORFs, they are involved in the generation of GDP-mannose and GDP-fucose, which are component sugars of colanic add. The E. coli CpsB and G proteins are homologous respectively to the CpsB and G and the RfbM and K proteins from the cps and rfb gene clusters of Salmonella enterica LT2 (2, 123). The high G+C ratio of the cloned cps genes from both E. coli and S. enterica LT2 suggests that the genes have been acquired from another organism with a high G+C DNA composition. The changes at the third base position suggest that the acquisition of these genes occurreed much earlier in E. coli than in S. enterica LT2, about 45 million years ago (2).

If, as is likely, group Ib K antigen gene clusters are located at the his region of the E. coli chromosome, then this region of the chromosome must contain the group Ib capsule gene cluster, the cps gene cluster for colanic acid, and the rfb gene cluster for the production of an O antigen together with a rol gene. Study of this region of the chromosome in group Ia and Ib strains will be fascinating and will help elucidate the relationships between these different group I K antigen gene clusters and their relationships to rfb gene clusters.

GROUP II K ANTIGENS

Group II K antigens have a higher charge density than those of group I and may contain hexuronic acids, NeuNAc, or KDO as acid components (62). They are coexpressed with many O antigens and are not expressed at growth temperatures below 20[degree]C (88). For certain group II K antigens, phosphatidic acid exists at the reducing end, and this may act as a membrane anchor that links the K antigen to the cell surface (59, 113). In contrast to E. coli strains that express group I K antigens, strains that express group II K antigens have elevated levels of the enzyme CMP-KDO synthetase at capsule permissive temperatures, and for certain group II K antigens, KDO attaches to the reducing end of the polysaccharide (35, 36). The genes that encode group II K antigens are termed kps and have been mapped near serA on the E. coli chromosome (90, 133).

The genes for a number of group II K antigen gene clusters have been cloned and analyzed. These include the genes for the K1 (118, 134), K4 (31), K5, K7, K12, and K92 antigens (104, 106).

Detailed molecular analysis of group II K antigen gene clusters revealed that different K antigen gene clusters have a conserved genetic organization that consists of three functional regions (Figure 1) (12, 13, 106). Regions 1 and 3 are conserved between different group II K antigen gene clusters, while the central region 2 is serotype specific (106, 107). Mutations within region 2 abolish polysaccharide biosynthesis, which suggests that this region encodes enzymes for the synthesis of particular K antigens (13, 106, 116). In the case of group II K antigens that contain sugars not normally ubiquitous in E. coli, additional enzymes involved in the biosynthesis of the appropriate nucleotide sugar precursors are also encoded within region 2. For instance, in the case of the K1 antigen, region 2 encodes three enzymes that are involved in the biosynthesis and activation of NeuNAc (132). Likewise, region 2 of the K5 capsule gene cluster encodes a UDP-Glc dehydrogenase enzyme for the synthesis of UDP-Glucuronic acid (GlcA), which is a component sugar of the K5 polysaccharide (97, 114). A correlation exists between the size of region

2 and the complexity of the K antigen. For instance, the K4 antigen, which is a complex branched heteropolymer (108), has a region 2 of 14 kb, in contrast to K1 region 2 of 5.8 kb (Figure 1) (11).

The best-studied group II capsule gene clusters to date are the K1 and K5 capsule gene clusters; the entire nucleotide sequence is now determined for the K5 capsule gene cluster (97, 103). The A+T ratio of the DNA within region 2 of the K5 capsule gene cluster was 66.6[percent] (97), which is higher than the 50[percent] A+T normally associated with E. coli chromosomal DNA (86). A similarly high A+T ratio has been reported for region 2 of the K1 capsule gene cluster (121).

Region 2 of the K5 capsule gene cluster is 8.0 kb and contains four genes with two large intergenic spaces (Figure 2) (97). Northern blot analysis and transcript mapping experiments identified three promoters within the K5 region 2 and revealed a complex pattern of transcription (97). Transcription from the kfiA promoter generates a transcript of 8.0 kb which spans the entire region 2, including the two large intergenic regions between the kfiA and B genes and kfiB and C genes (Figure 2). Two smaller transcripts that originate from the intergenic regions between kfiA and B and kfiB and C genes generate transcripts of 6.0 and 3.1 kb. (Figure 2) (97). Region 2 of the K1 capsule gene cluster is 5.8 kb and contains six genes, all of which are transcribed in the same direction as region 3 (1, 115, 132).

The serotype-specific region 2 is flanked by regions 1 and 3 (Figures 1 and 2). Both regions 1 and 3 are conserved between different group II K antigen gene clusters and encode common functions (103). The mechanism of acquisition of different forms of region 2 at this site is discussed below. The complete nucleotide sequence has been determined for the entire region 1 of the K5 capsule gene cluster (93, 94) and partially determined for region 1 of the K1 capsule gene cluster (20, 141). The A+T ratio of the DNA was 50.6[percent], a value typical for E. coli (93). Region 1 contains six genes, kpsFEDUCS (Figure 2), two of which, kpsU and kpsC, appear to be translationally coupled (93, 94). Translational coupling is a proposed mechanism by which balanced expression of two proteins may be achieved (144), which may be of particular significance if the two proteins interact in some form of complex.

Northern blotting and transcript mapping confirm that region 1 is organized in a single transcriptional unit with the promoter located 225 bp upstream from the initiation codon for the first gene kpsF (Figure 2) (D Simpson & I Roberts, unpublished data). Analysis of the promoter revealed no similarities with promoters recognized by alternative sigma factors. Transcription from the region 1 promoter was temperature regulated, and there is no detectable transcription at 18[degree]C, a nonpermissive temperature for capsule expression.

Region 3 of group II capsule gene clusters contains two genes kpsM and T that are organized as a single transcriptional unit (Figure 2) with both genes translationally coupled (92, 119). This coupling suggests, as mentioned above, that the KpsM and T proteins are likely to interact in some form of complex. The promoter for the K5 region 3 is 741 bp upstream of the initiation codon for KpsM and has no similarities with promoters recognized by alternative sigma factors in E. coli (M Stevens & I Roberts, unpublished data). Because of the high degree of identity between region 3 from the K1 and K5 capsule gene clusters, the promoter is probably located at a similar site in the K1 capsule gene cluster.

Transcription of region 3 is in the same direction as transcription in region 2 in both the K1 and K5 capsule gene clusters (97, 115). Transcription from the region 3 promoter was temperature regulated in a manner analogous to that of the region 1 promoter (M Stevens & I Roberts, unpublished data).

GROUP III K ANTIGENS

Recently, the genes for a third group of E. coli capsule gene clusters have been cloned (96). This group, typified by the K10 and K54 antigens, is encoded by genes that map to the same serA site on the chromosome as do group II capsule genes (89, 90). Preliminary genetic analysis of the K10 and K54 capsule gene clusters (96) suggests that group III capsule gene clusters have a conserved genetic organization that contains a central serotype-specific region flanked by group III capsule-specific sequences in a manner analogous to that of group II capsule gene clusters (96). However, the group III capsule genes appear to have little detectable

nucleotide sequence in common with the group II capsule genes, which suggests that these capsule gene clusters may have originated from a different source than that of the group II capsule genes, but have been inserted at the same serA site on the E. coli chromosome (96).

ANALYSIS OF CAPSULE GENE CLUSTERS FROM OTHER GRAM-NEGATIVE BACTERIA

The genes have been cloned for a number of both group I-like and group II-like capsules from a range of gram-negative bacteria. The genes for the production of group I-like capsules have been cloned from E. amylovora, Erwinia stewartii, P. solanacearum, and K. pneumoniae (3, 18, 29, 57). In these cases the genes are clustered. Both the E. amylovora and P. solanacearum capsule gene clusters appear to be transcribed as single large transcripts, although other minor promoters may be present within the gene clusters and their roles cannot be ruled out (18, 57).

The genes for the production of group II-like capsules from Haemophilus influenzae type b, Neisseria meningitidis group B, and the Salmonella typhi Vi antigen have been cloned (42, 53, 70). The organization of these capsule gene clusters is in many ways remarkably similar to that of the group II capsule gene clusters of E. coli. The H. influenzae and N. meningitidis capsule gene clusters are aligned with the E. coli K5 cluster (Figure 2).

The H. influenzae type b capsule gene (cap) cluster has a conserved genetic organization (Figure 2). A central serotype-specific region 2 encodes the polysaccharide biosynthetic functions (128) and is flanked by regions 1 and 3, which are common to all of the H. influenzae serotypes (68, 70). Region 1 of the type b cap cluster contains four genes bexABCD that are probably organized as a single transcriptional unit (68). The nucleotide sequence of region 2 contains four genes, one of which encodes CDP-ribitolpyrophosphorylase, an enzyme required for the biosynthesis of the type b polysaccharide (128). The cap locus is duplicated in most type b strains: Two directly repeated copies of 17 kb are linked by a bridge region that contains the single unique copy of the bexA gene, which is the last gene in the cap gene cluster (69, 67). The mechanisms by which the H. influenzae cap locus may be lost or its copy number increased are discussed below.

The capsule gene cluster (cps) from N. meningitidis group B has a complex segmental organization (Figure 2). Originally, five regions were identified within the cps gene cluster (42). Region A was thought to be involved in polysaccharide biosynthesis, regions C and D in polysaccharide transport, and regions D and E in regulation of capsule expression (42). Further experimental work has redefined this preliminary hypothesis. Region C contains four genes ctrABCD that are most likely organized in a single transcriptional unit (39) while region B contains two genes termed lipAB (Figure 2), originally thought to be organized as two transcriptional units (40). However, careful analysis of the nucleotide sequence indicates that the lipA gene is larger than first thought, and therefore the two genes are likely to be transcribed as a single unit (I Roberts, unpublished results). Four genes were identified within region A (Figure 2) that code for proteins involved in the biosynthesis of the group B polysaccharide (33, 43). Further analysis of the nucleotide sequence of region D allowed identification of the galE gene and the three genes rfbBCD (51). The rfbBCD genes encode proteins homologous to those involved in the biosynthesis of rhamnose, which contains LPS molecules in Salmonella typhimurium (51). The absence of any detectable rhamnose-containing polysaccharides in meningococci and the incomplete expression of all of the rfb genes are puzzling. When the galE gene was expressed, mutations in the gene resulted in truncated lipooligosaccharide (LOS) that was not sialylated (51). The lack of any LOS or LPS genes within the H. influenzae and E. coli capsule gene clusters suggests that, in the case of the N. meningitidis group B capsule gene cluster, the insertion of the LOS genes occurs by some form of chromosomal rearrangement within meningococci.

THE BIOCHEMISTRY OF CAPSULE PRODUCTION IN GRAM-NEGATIVE BACTERIA

The biochemistry of E. coli group II capsule production has been studied exhaustively (35, 62, 126, 139). A combination of both biochemical and genetic techniques has revealed much about the mechanisms by which group II polysaccharides are synthesized and exported in E. coli (103).

THE EXPORT OF GROUP II POLYSACCHARIDES

Although fundamental differences may exist between the export of group II-like capsules in E. coli, H. influenzae, and N. meningitidis, uniform themes are conserved in the expression of polysaccharide capsules in these taxonomically diverse bacteria. It is unclear whether this reflects a common ancestry of capsule biosynthesis genes (41) or is instead the result of underlying chemical constraints that limit the diversity of the process. As a consequence, much of what has been learned about the export of group II capsules in E. coli can be extrapolated to capsule expression in these other bacteria.

The functions performed by the proteins encoded by regions 1 and 3 in the export of group II capsular polysaccharides in E. coli are only partially elucidated. Mutations within region 3 result in cytoplasmic polysaccharide associated with the inner face of the cytoplasmic membrane (13, 71, 91, 99, 119). This suggests that the proteins encoded by region 3 are involved in the export of group II polysaccharides across the cytoplasmic membrane, and it confirms that polysaccharide biosynthesis occurs on the inner face of the cytoplasmic membrane. Analysis of the predicted amino acid sequence of the KpsM and T proteins indicates that they are ATP-binding cassette (**ABC**) **transporters** and may comprise an inner-membrane polysaccharide export system (91, 92, 119).

ABC transporters are ubiquitous, and they are involved in a diverse range of import and export systems in both prokaryotes and eukaryotes (28). These include bacterial transport systems for the uptake of oligopeptides and amino acids (54) as well as for the export of polysaccharides (34). Members of the ABC superfamily have a common organization that consists of a hydrophobic integral membrane protein and a hydrophilic-membrane associated ATP-binding protein (28). The transport complex consists of two of each component, and in some cases the different components are fused into a single multi-domain protein (54).

In the case of E. coli group II polysaccharide export, the KpsT protein has an ATP-binding fold, and a hydrophobicity plot of the predicted amino acid sequence of KpsM reveals a protein with at least six potential membrane-spanning domains (92, 119). Site-directed mutagenesis and binding of azidolinked ATP have confirmed the role of the KpsT protein as an ATP-hydrolyzing protein (91, 115, 117). Structure-function studies on the KpsM protein confirm that its membrane topology has six membrane-spanning domains with two cytoplasmic and three periplasmic loops (99, 115). Therefore, the group II polysaccharide transport complex is likely to consist of two molecules of KpsM that may form some type of inner membrane-spanning pore, associated with two molecules of KpsT that catalyze ATP and energize the transport process. Mutations in region 3 of the K1 capsule gene cluster could be complemented in trans by the cloned region 3 of the K5 capsule genes, which indicates that the KpsM and T proteins are able to export group B polysaccharides that are independent of the repeat structure of the polysaccharide (106).

The homology of the KpsM and T proteins and the proteins that are encoded by the H. influenzae cap cluster and N. meningitidis cps cluster (Table 2) suggests that **ABC transport** systems operate to export capsular polysaccharides across the cytoplasmic membrane in these bacteria (39, 69). Mutations in region 3 of the K5 capsule gene cluster are complemented by cloned genes from Actinobacillus pleuropneumoniae (137), which confirms the functional conservation that appears to exist in the **ABC transport** systems involved in the export of group II polysaccharides in gram-negative bacteria.

The functional conservation in the export of capsular polysaccharides in gram-negative bacteria has been confirmed by computer-aided database searches. These searches of the predicted amino acid sequence of the six proteins that are encoded by region 1 revealed similarities to proteins that are involved in the expression of polysaccharide capsules in other bacteria (Table 2). The KpsF protein has no homologue in other capsule gene clusters but has significant similarity to the GutQ protein (Table 2), a hypothetical protein ORF328 of E. coli, and a KpsF homologue in H. influenzae (D Simpson & I Roberts, unpublished data). The GutQ protein may act as a regulator of the glucitol operon in E. coli (143); thus, the KpsF protein may have a regulatory role in the expression of group II capsules in E. coli (20). However, mutants lacking a functional kpsF gene are still

able to express a K5 capsule, and they demonstrate temperature-dependent capsule expression (14, 93, 94). In order to preclude the possibility that the mutation in the kpsF gene was complemented in trans by the gutQ gene, the experiments were repeated in a gutQ mutant. In this case, a temperature-dependent expression of capsule genes was still observed (D Simpson & I Roberts, unpublished data). Therefore, the role of the KpsF protein is still unknown, but the high degree of identity (95[percent]) between the KpsF proteins encoded by the K1 and K5 capsule gene clusters would suggest some functional role for the KpsF protein. The situation is not made any clearer by the identification of a KpsF homologue in H. influenzae because the corresponding kpsF gene is not located with the H. influenzae cap gene cluster (D Simpson & I Roberts, unpublished data).

Mutations within the kpsE and kpsD genes result in periplasmic polysaccharide (13, 14), which suggests a role for these two proteins in the export of polysaccharide across the periplasmic space. The KpsE protein is homologous to the BexC protein of H. influenzae and the CtrB protein of N. meningitidis (Table 2), both of which are implicated in the export of capsular polysaccharides in these two bacteria (39, 68). Rosenow et al have purified the KpsE protein encoded by the K5 capsule gene cluster and generated antisera to the protein (110). Both biochemical and genetic techniques allow elucidation of the topology of the KpsE protein within the inner-membrane (110). Analysis of KpsE-BlaM fusions demonstrates that the KpsE protein has a cytoplasmic N-terminus with a large periplasmic domain of approximately 300 amino acids and a C-terminal membrane-associated domain that is unlikely to extend across the inner-membrane into the cytoplasm (F Esumeh & I Roberts, unpublished data). This arrangement of the KpsE protein within the inner-membrane suggests that the periplasmic domain of the KpsE protein may be functionally important in the export of group II polysaccharides. Based on predicted amino acid sequence, both the BexC and CtrB proteins are likely to have similar topologies within the inner-membrane.

The KpsD protein is a periplasmic protein with a typical N-terminal signal sequence (93, 94, 141). Mutations in the kpsD gene result in periplasmic polysaccharides, which suggests a role for KpsD in the passage of group II polysaccharides across the periplasmic space following export across the cytoplasmic membrane (14, 141). The KpsD protein likely interacts with the periplasmic domain of the KpsE protein when mediating the export of group II polysaccharides across the periplasmic space. However, a homologue to the KpsD protein is not encoded by either the H. influenzae or N. meningitidis capsule gene clusters. The KpsD protein is homologous to the ExoF protein (Table 2), which is implicated in the expression of succinoglycan in R. meliloti (83). In the case of E. coli group II capsular polysaccharides, the means of export of polysaccharide across the outer membrane is not clear. No outer-membrane protein is encoded by the kps gene cluster, unlike the capsule gene clusters of H. influenzae and N. meningitidis, which both encode an outer-membrane protein that may play a role in the final stages of capsular polysaccharide export (41, 68). The lack of an outer-membrane protein encoded by the kps gene cluster means either that another outer-membrane protein may perform this function, or that export of capsular polysaccharide in E. coli onto the cell surface requires no outer-membrane protein and is achieved in a fashion different from that in H. influenzae and N. meningitidis. E. coli capsular polysaccharide export onto the cell surface may be achieved by the formation of membrane fusions or Bayer sites between the inner and outer membrane (103). While this suggestion is appealing because of the possible cycling of phospholipid linked polysaccharide through the fused membranes, little experimental evidence exists at this stage to support this notion.

The KpsC and KpsS proteins are located in the cytoplasm associated with the inner face of the cytoplasmic membrane (G Rigg & I Roberts). Mutations within either the kpsC or kpsS genes result in aggregates of polysaccharide that accumulate within the cytoplasm and lack both phospholipid and KDO, which are normally found at the reducing end of cell-surface polysaccharide (14, 15). Therefore the KpsC and KpsS proteins may be involved in the attachment of KDO to phospholipid and the subsequent ligation of the polysaccharide to the phosphatidyl-KDO prior to export across the cytoplasmic membrane by KpsM and KpsT (103). Such an interpretation suggests that the presence of phosphatidyl-KDO at the reducing end of the polysaccharide molecule is the motif recognized by the export proteins, which might explain how a conserved set of proteins could

export a broad range of chemically diverse group II polysaccharides. Indeed, a similar lipidation function has been suggested for the LipA and LipB proteins encoded by the N. meningitidis capsule gene cluster (40), which are homologous to KpsC and KpsS respectively (Table 2). However, to date, no biochemical evidence for the function of either KpsC or KpsS proteins is available. Furthermore, KDO has not been demonstrated to be at the reducing terminus of all group II polysaccharides in E. coli.

The KpsC protein is also homologous to LpsZ (Table 2) (93). This protein is responsible for modification of LPS molecules in R. meliloti (17). The functional significance of this similarity is unclear, especially in conjunction with evidence that the cloned lpsZ gene does not complement a kpsC mutant (I Roberts, unpublished data).

The KpsU protein is 44[percent] identical to the KdsB enzyme of E. coli (93, 94). The KdsB enzyme is the CMP-KDO synthetase enzyme that catalyzes the formation of CMP-KDO prior to the linkage of KDO to lipid A (46). Rosenow et al purified the KpsU protein and demonstrated it to be a CMP-KDO synthetase enzyme (109), which explains the elevated levels of this enzyme in E. coli strains expressing group II capsules (36). The KpsU enzyme probably generates CMP-KDO, which is a substrate for the attachment of KDO to phospholipid (103). Mutations in the kpsU gene of the K5 capsule gene cluster do not abolish capsule production (14), most likely owing to complementation of the defect by the kdsB gene.

Mutations in any of the genes in regions 1 or 3 result in a significant reduction in the membrane transferase activity (15, 91). Such a disruption of polysaccharide export affects polysaccharide biosynthesis, indicating that the two are linked. The proteins involved in polysaccharide synthesis and export presumably form a multi-protein complex on and through the inner-membrane with many protein-protein interactions (103). A multi-protein complex for the E. coli K5 antigen does exist on the inner-membrane, and mutations in particular region 1 genes have pronounced pleiotropic effects on the assembly of this complex (G Rigg & I Roberts, unpublished data).

The processes involved in the export of group II polysaccharides in E. coli are beginning to be clarified by research and by extrapolation in other gramnegative bacteria. Despite common themes, export processes may differ significantly: for instance, in the apparent lack of an outer-membrane protein in the transport process in E. coli, in the role of KDO in the transport process, and in the lack of an obvious capsule-associated periplasmic protein in both H. influenzae and N. meningitidis.

GROUP II POLYSACCHARIDE BIOSYNTHESIS

The biosynthesis of the E. coli K1 and K5 antigens and the group B capsule of N. meningitidis have been extensively studied (35, 62, 78, 79, 121, 126, 139). In the case of the E. coli K1 and K5 antigens, the polysaccharide grows at its nonreducing end by the sequential addition of monosaccharide units (35, 120, 139). The biosynthesis of NeuNAc is achieved by the conversion of N-acetylmannosamine-6-phosphate to N-acetylmannosamine (ManNAc), which is probably catalyzed by the NeuC protein (136). The NeuB protein then catalyzes the condensation of ManNAc with phosphoenolpyruvate to generate NeuNAc (1), which is activated to form CMP-NeuNAc by the NeuA enzyme (145). The polymerization of the K1 antigen is then achieved by NeuS, a processive polysialyltransferase enzyme (120). Functions for the NeuD and NeuE proteins encoded by region 2 of the K1 capsule gene cluster are less well known. Mutations in the neuD gene abolish K1 production, and NeuD possibly modifies other proteins involved in K1 biosynthesis (1). Mutations in the neuE gene result in intracellular K1 polysaccharide (135). NeuE may play a role in the assembly of the polymerization-export complex, and the ability of NeuE to bind to undecaprenol phosphate may be essential for this process (126). Whatever the function of NeuE, it is likely to be specific to E. coli group II capsules that contain NeuNAc.

The steps involved in the biosynthesis of the group B capsule in N. meningitidis are similar to the steps of K1, with one or two key differences. The biosynthesis of NeuNAc and its activation to CMP-NeuNAc appear to proceed via an identical pathway, and significant similarities exist between the meningococcal and E. coli NeuA, NeuB, and NeuC proteins (43). However, no homologues of either the E. coli NeuE or NeuD proteins have been identified in N. meningitidis group B (33). In addition, the polysialyltransferases are 30[percent] identical (39), but the

meningococcal polysialyltransferase appears to be the only enzyme required for the initiation and elongation of polysaccharide biosynthesis of the group B polysaccharide (33). Different enzymes in N. meningitidis group B and E. coli K1 regulate the availability of NeuNAc. In meningococci, a CMP-NeuNAc hydrolase cleaves CMP-NeuNAc (78). The gene that encodes this enzyme is not present in the group B capsule gene cluster but maps elsewhere on the meningococcal chromosome (33). E. coli K1 lacks this enzyme activity, and the levels of NeuNAc are regulated by a NeuNAc aldolase, an enzyme that is also not encoded within the K1 capsule gene cluster (136).

The biosynthesis of the E. coli K5 antigen requires four proteins, KfiA-D, encoded within region 2 (Figure 1) (97). The functions of two of the encoded proteins were confirmed by over-expression of individual genes and subsequent assays of the recombinant proteins (97). The KfiC protein is a bifunctional glycosyltransferase enzyme that adds alternating glucuronic acid and N-acetylglucosamine residues at the nonreducing end of the growing polysaccharide chain (97). The KfiC enzyme is unable to initiate K5 biosynthesis itself, which suggests that the initial reactions for K5 biosynthesis are mediated by other enzymes encoded within region 2. Western blot analysis with **antibodies** to the purified KfiC enzyme confirms the enzyme's location on the inner face of the cytoplasmic membrane and reveals that association of the KfiC enzyme with the cytoplasmic membrane requires proteins encoded by regions 1 and 3 (G Rigg, G Griffiths, & I Roberts, unpublished data). This confirms the notion that a polysaccharide biosynthetic complex exists on the inner face of the cytoplasmic membrane. Alignment of the predicted amino acid sequence of KfiC with other known glycosyltransferase enzymes reveals three conserved regions that are likely to be functionally important in projective glycosyltransferase enzymes (97). The KfiD protein has been demonstrated to be a UDP-glucose dehydrogenase that converts UDP-glucose to UDP glucuronic acid, which is a component sugar of the K5 polysaccharide (114).

The functions of the KfiA and B proteins are so far unknown. Preliminary evidence suggests a role for these enzymes in the initial stages of K5 biosynthesis: possibly in the attachment of the first sugar residue to a membrane acceptor onto which the K5 polysaccharide is then synthesized by the KfiC enzyme (103). The nature of the membrane acceptor for the initiation reaction is still unknown. Whatever the function of both the KfiA and B proteins, they appear to be associated with the polysaccharide biosynthetic complex on the inner face of the cytoplasmic membrane (G Rigg & I Roberts, unpublished data).

BIOSYNTHESIS OF GROUP I POLYSACCHARIDES

The biochemistry of group I polysaccharide biosynthesis in E. coli is less well understood. By extrapolation from the biosynthesis of group I-like capsules from other bacteria, the biosynthesis is considered to involve two steps: first, the generation of undecaprenol-linked intermediates, and second, the polymerization of these intermediates at the reducing terminus to form polysaccharide.

The biosynthesis of the capsular polysaccharide of Aerobacter aerogenes strain DD45 has been studied in detail (127). In this case, the tetrasaccharide repeat unit of the polysaccharide is synthesized on an undecaprenol phosphate carrier. Polymerization then occurs by the formation of a glycosidic bond and the transfer of one repeat unit to another, which is itself linked to undecaprenol phosphate, thereby forming an undecaprenol phosphate-linked molecule with two repeat units attached. The freed undecaprenol pyrophosphate is then dephosphorylated, which allows a cycling of the undecaprenol phosphate (127). This dephosphorylation reaction is inhibited by bacitracin (124), and the sensitivity of capsular polysaccharide biosynthesis to bacitracin indicates a role for the cycling of undecaprenol phosphate in capsule production. The biosynthesis of a number of extracellular polysaccharide molecules, including xanthan gum of Xanthomonas campestris (58), also proceeds via lipid linked oligosaccharides.

Colanic acid biosynthesis in E. coli is believed to be achieved by a similar mechanism using undecaprenol phosphate (140). The similarity between E. coli group Ia capsules and the capsules of Klebsiella suggests that group Ia capsules are synthesized in the same fashion. That group Ib capsules may be regarded as acidic O antigens (27, 59) suggests the

following biosynthetic process: Group Ib polysaccharide polymerization might occur at the reducing terminus using blocks of undecaprenol-linked repeating units, as observed with heteropolymexic O antigens (140). In the next few years, the biosynthesis of group I capsules should become clearer.

THE GENETICS OF CAPSULE PRODUCTION IN GRAM-POSITIVE BACTERIA

Capsule gene clusters have been cloned from a number of gram-positive bacteria, including Streptococcus pneumoniae (26, 44, 49), Staphylococcus aureus (73, 75), and group B streptococci (111). Because the biochemistry and export of these polysaccharides are as yet only poorly understood, I focus on the genetics of capsule production in these organisms.

The gene clusters for the S. pneumoniae type 3 and 19F capsules have been cloned and analyzed (26, 44, 49). To date, seven genes have been identified within the type 3 capsule gene cluster (19, 25, 26). Four of these genes, cps3DSUM, are specific for the type 3 capsule. The genes are organized into two transcriptional units, cps3DS and cps3UM (25). This capsule-specific region is flanked by two regions common to pneumococci that express other capsule serotypes. Based on predicted amino acid sequence homology, possible functions were assigned to the type-specific proteins: Cps3D was assigned a UDP glucose dehydrogenase; Cps3S, a polysaccharide synthase; Cps3U, a glucose-1 phosphate uridylytransferase; and Cps3M, a phosphoglucomutase (25). Mutations within the cps3U or cps3M genes do not abolish capsule production, which suggests that the functions performed by these proteins can be complemented by other enzymes in the cell (19). Immediately upstream of the cps3D gene is a 938-bp segment that encodes a small ORF and contains sequences that are repeated in the chromosomes of S. pneumoniae strains, expressing types 2, 3, 5, 6B, 8, and 22 (19). This region is not thought to play a role in the expression of the type 3 capsule but may be important in the acquisition of capsule gene clusters in S. pneumoniae (25). Three other genes, cps3BCP, have been identified upstream of the repeated sequence and are homologous to the cpsfBCD genes of the type 19F capsule gene cluster (19). Downstream of the cps3M gene are common sequences that in the case of the type 3 capsule contain the two genes tnpA and plpA (19, 25). However, both of these genes have undergone deletions and encode nonfunctional proteins (19). The TnpA protein is homologous to a number of transposases, but it lacks 200 amino acids from its N-terminal end and 490 amino acids from its C-terminus. The PlpA protein is a permease-like protein that is important in transformation in S. pneumoniae (95). In the case of the plpA gene that flanks the type 3 capsule genes, the PlpA protein lacks the first 281 amino acids and is therefore defective. Polymerase chain reaction (PCR) has been used to analyze the sequences downstream of the type 3 capsule gene clusters in unrelated pneumococci that express type 3 capsules. Such analysis has made it possible to confirm the presence of defective tnpA and plpA genes downstream of all type 3 capsule gene clusters, which indicates that all type 3 strains are plpA minus (19). Recent data suggest that both CpsM and PlpA may play important roles in sensing the environment and causing up-regulation of capsule expression in S. pneumoniae (J Yother, unpublished results). The PlpA system may operate in non-type 3-strains, and the Cps3M-mediated pathway may be in operation in type 3 strains that lack a functional PlpA (J Yother, unpublished results).

Analysis of the type 19F capsule gene cluster revealed the presence of seven genes, cps19FA-G, organized in a single transcriptional unit (49). The Cps19FA protein is homologous to the transcriptional activator LytR, which suggests a possible role for this protein in the regulation of capsule expression. The cpsl9FA gene appears to be conserved in all of the serotypes analyzed, which suggests a general role for this protein in pneumococcal capsule expression (49). The function of the Cps19FB protein is not clear, but the protein is conserved in all of the serotypes analyzed. The Cps19FC and D proteins are homologous to the Cps3C and P proteins, and these proteins may be involved in the export of pneumococcal capsular polysaccharides (49). Comparison of the gene clusters for the type 3 and 19F capsules shows that conserved genes exist between the different capsule gene clusters, and some form of genetic organization is conserved. However, the presence of additional DNA in the type 3 capsule gene cluster and differences in the transcriptional organization in genes that are conserved between different serotypes suggest that the conditions

in S. pneumoniae may be more complex than in E. coli. This complexity may reflect the ease by which genetic information can be taken up and exchanged within pneumococci.

The genes for the production of both the type 1 and type 5 capsule serotypes have been cloned from Staphylococcus aureus (73, 75). The type 1 capsule gene cluster is located on a large discrete genetic element of approximately 34 kb and appears specific to strains that express the type 1 capsule (74). This is in contrast to the type 5 capsule gene cluster. Southern blot experiments using probes that spanned the type 5 capsule gene cluster identified sequences common to all capsule types, as well as capsule-specific sequences reminiscent of group II capsule gene clusters in E. coli (75). Nucleotide sequence data, when they become available, should resolve the apparent differences between the organization of the type 1 and type 5 capsule gene clusters of S. aureus.

THE GENERATION OF CAPSULE DIVERSITY

The huge number of chemically different capsular polysaccharides prompts the question: How has this diversity been achieved? In the case of gram-negative bacteria, the high A+T ratio of the DNA that encodes the polysaccharide biosynthesis enzymes (33, 43, 97, 121, 128) suggests a common ancestry of these capsule biosynthesis genes (41). In group II E. coli capsule gene clusters, the high A+T ratio of region 2 DNA, as compared to regions 1 and 3, would confirm that group II capsule diversity has been achieved in part through the acquisition of different region 2 sequences. The lack of any conserved sequences of different group II capsule gene clusters between regions 1 and 2 and regions 2 and 3 may preclude some form of site-specific transposition event as a means of changing region 2 sequences (30). Rather, the acquisition of new region 2 sequences may occur through homologous recombination of an incoming and resident capsule gene cluster between the flanking regions 1 and 3. The marked divergence of the C-termini of the KpsS and KpsT proteins from different group II capsule gene clusters (30, 91, 97) would support this hypothesis. The 3' ends of the kpsS and kpsT genes are located, respectively, adjacent to the junctions of regions 1 and 2 and regions 2 and 3; therefore, these differences in the C-termini of the proteins may have arisen as a result of recombination events between regions 1 and 3 of different capsule gene clusters. The mechanism by which group II capsule genes have been acquired and why this set of different capsule genes is inserted at the same serA region of the chromosome are unknown.

The acquisition of regions 2 by homologous recombination may also be responsible for the generation of capsule diversity in both H. influenzae and N. meningitidis. However, in both these bacteria, additional chromosomal rearrangements are likely to be important in the expression of capsular polysaccharides. In N. meningitidis group B, expression of group B polysaccharide can be switched on and off by the insertion/excision of sequence IS1301 in the neuA gene (52). The inactivation of the neuA gene abolishes the production of CMP-NeuNAc, thereby abolishing both capsule production and sialylation of LOS. Loss of capsule will likely promote adherence and invasion of epithelial cells (52). The subsequent restoration of capsule production and LOS sialylation by the spontaneous excision of IS1301 from the neuA gene would permit the survival of the invading meningococci, by conferring resistance to complement-mediated killing and phagocytosis.

In H. influenzae type b, the duplicated cap locus lies between direct repeats of a IS1016 and essentially generates a compound transposon that contains the capsule gene cluster (69). The presence of the IS1016 elements allows the amplification of the cap locus, thereby augmenting type b capsule production. At one end of the duplicated cap gene cluster, there is a 1.2-kb deletion that removes most of one copy of the bexA gene. The remaining functional bexA gene is located in the bridge region that links the two copies of the cap gene cluster so that directly repeated capsule genes are necessary for capsule expression (67, 69). This arrangement of the cap locus is preserved among nearly all H. influenzae type b strains responsible for the vast majority of invasive Haemophilus infections worldwide. Such an observation suggests that this deletion and cap gene duplication occurred in an ancestral type b strain and generated some form of selective advantage (69).

In S. pneumoniae, capsule type changes as a consequence of

transformation with donor DNA from a different capsule type. Homologous recombination between conserved sequences that flank the serotype-specific region serves as a likely model. As a rare consequence of transformation, two capsule types (binary encapsulation) may be produced by a single transformant (4). In the case of unstable binary encapsulation, the second capsule gene cluster is linked to the first, suggesting some form of illegitimate recombination event in the flanking sequences that is eventually resolved with the end result of a single capsule type (25). Stable binary transformants have the second capsule gene cluster inserted at a second distal site on the chromosome (8). Although a number of possibilities exist, the basis for this stable acquisition of a second capsule gene cluster is not known. Hopefully, with a greater understanding of the molecular genetics of capsule gene clusters in S. pneumoniae, it may be possible to address this question.

PERSPECTIVES

Bacterial capsular polysaccharides are a diverse range of biologically important molecules. They play pivotal roles in mediating a number of biological processes, particularly in affecting microbe-host interactions during the onset and development of infectious disease. The use of molecular genetic techniques has moved forward our understanding of the organization of capsule gene clusters in a wide variety of bacteria and begun to shed light on the process of capsule diversity. The comparison of predicted amino acid sequences of encoded proteins illustrates that conserved themes, such as **ABC** -polysaccharide **transport** systems, are present in diverse bacteria. However, there are still huge areas for which at the molecular level our knowledge is at best slight. This is particularly true for the export of polysaccharides in gram-negative bacteria. It seems to me that we have to think of novel biochemical approaches to begin to answer these questions. The prize is not only the knowledge itself; rather, it is the production of chemotherapeutic agents targeted to selectively disrupt capsule export and thus combat infections by encapsulated bacteria. The extension of our studies on capsule production to eukaryotic pathogens of human beings promises to be exciting.

Added material

Ian S. Roberts
School of Biological Siences, University of Manchester, Manchester M13 9PT, United Kingdom

Table 1 Functions of polysaccharide capsules

Table 2 Homology between proteins encoded by regions 1 and 3 of the K5 capsule gene cluster

Protein	Cellular location	Similarity
KpsF	Cytoplasm	GutQ (72[percent] over 314 aa)
KpsE	Inner membrane	BexC (73[percent] over 359 aa)
		CtrB (73[percent] over 355 aa)
KpsD	Periplasm	ExoF (67[percent] over 100 aa)
KpsU	Cytoplasm	KdsB (70[percent] over 246 aa)
KpsC	Associated with the inner face of the cytoplasmic membrane	LpsZ (76[percent] over 312 aa)
		LipA (70[percent] over 550 aa)
KpsS	Cytoplasm	LipB (68[percent] over 396 aa)
		NeuA (66[percent] over 246 aa)
KpsM	Inner membrane	BexB (69[percent] over 266 aa)
		CtrC (68[percent] over 266 aa)
KpsT	Associated with the inner face of the cytoplasmic membrane	BexA (89.4[percent] over 220 aa)
		CtrD (85[percent] over 220 aa)

Figure 1 Schematic representation of the organization of E. coli group II capsule gene clusters. The K1 capsule gene cluster is shown with the three functional regions. The boxes labeled K92, K5, and K4 represent the serotype-specific region 2s that are inserted between the conserved regions 1 and 3.

Figure 2 Genetic organization of the E. coli K5, H. influenzae type b, and N. meningitidis group B capsule gene clusters. For clarity, a single copy of the H. influenzae capsule gene cluster is shown. The large boxes denote the conserved functional regions (see text). The smaller labeled boxes define specific genes that have been identified within each cluster.

The hatched boxes in region 2 of the K5 capsule gene cluster define intergenic gaps.

ACKNOWLEDGEMENTS

I thank the large number of people who have worked over the years in my laboratory, as well as my collaborators. I owe them a great debt for their hard work and enthusiasm. In particular I thank Professors Bob and Barbara Jann. In addition, I appreciate the help of Drs. M Frosch, C Y Lee, C Rubens, E Vimr, C Whitfield, and J Yother, all of whom sent me information and were happy to discuss their work. Thanks to Gillian Etches for her excellent secretarial help. The work in my laboratory is supported by the Biotechnology and Biological Sciences Research Council and Medical Research Council of the United Kingdom, together with the Wellcome Trust. I gratefully acknowledge the support of the Lister Institute of **Preventive** Medicine.

LITERATURE CITED

1. Annunziato P, Wright L, Vann W, Silver R. 1995. Nucleotide sequence and genetic analysis of the neuD and neuB genes in region 2 of the polysialic acid gene cluster of Escherichia coli K1. J. Bacteriol. 177:312-19

2. Aoyama K, Haase AM, Reeves PR. 1994. Evidence for effect of random genetic drift G+C content after lateral transfer of fucose pathway genes to Escherichia coli K-12. Mol. Biol. Evol. 11(6):829-38

3. Arakawa Y, Wacharotayankun R, Nagatsuka T, Ito H, Kato N, Ohta M. 1995. Genomic organisation of the Klebsiella pneumoniae cps region responsible for serotype K2 capsular polysaccharide synthesis in the virulent strain Chedid. J. Bacteriol. 177:1788-96

4. Austrian R, Bernheimer HP, Smith EEB, Mills GT. 1959. Simultaneous production of two capsular polysaccharides by pneumonococcus. II. The genetic and biochemical bases of binary capsulation. J. Exp. Med. 110:585-602

5. Bastin DA, Stevenson G, Brown PK, Haase A, Reeves PR. 1993. Repeat unit polysaccharides of bacteria: a model for polymerization resembling that of ribosomes and fatty synthetase, with a novel mechanism for determining chain length. Mol. Microbiol. 7:725-34

6. Batchelor R, Alifano P, Biffali E, Hull S, Hull A. 1992. Nucleotide sequences of the genes regulating 0-polysaccharide chain length (rol) from Escherichia coli and Salmonella typhimurium: protein homology and functional complementation. J. Bacteriol. 174:5228-36

7. Bayer M, Thurow H. 1977. Polysaccharide capsule of Escherichia coli: microscope study of its size, structure, and sites of synthesis. J. Bacteriol. 130:911-36

8. Bernheimer HP, Wermundsen IE, Austrian R. 1968. Mutation in Pneumococcus type III affecting multiple cistrons concerned with the synthesis of capsular polysaccharide. J. Bacteriol. 6:1099-2002

9. Berry A, DeVault J, Chakrabarty A. 1989. High osmolarity is a signal for enhanced algD transcription in mucoid and nonmucoid Pseudomonas aeruginosa strains. J. Bacteriol. 171:2312-17

10. Bhattacharjee A, Jennings H, Kenny C, Martin A, Smith I. 1975. Structural determination of the sialic acid polysaccharide antigens of Neisseria meningitidis serogroups B and C with carbon 13 nuclear magnetic resonance. J. Biol. Chem. 250:1926-32

11. Boulnois G, Drake R, Pearce R, Roberts I. 1992. Genome diversity at the serA-linked capsule locus in Escherichia coli. FEMS Microbiol. Lett. 100:121-24

12. Boulnois GJ, Roberts IS. 1990. Genetics of capsular polysaccharide production in bacteria. Curr. Top. Microbiol. Immunol. 150:1-18

13. Boulnois GJ, Roberts IS, Hodge R, Hardy KR, Jann KB, Timmis KN. 1987. Analysis of the K1 capsule biosynthesis genes of Escherichia coli: definition of three functional regions for capsule production. Mol. Gen. Genet. 208:242-46

14. Bronner D, Sieberth V, Pazzani C, Roberts IS, Boulnois GJ, et al. 1993. Expression of the capsular K5 polysaccharide of Escherichia coli: biochemical and electron microscopic analyses of mutants with defects in region 1 of the K5 gene cluster. J. Bacteriol. 175:5984-92

15. Bronner D, Sieberth V, Pazzani C, Smith A, Boulnois G, et al. 1993. Synthesis of the K5 (group II) capsular polysaccharide in transport-deficient recombinant Escherichia coli. FEMS Microbiol. Lett. 113:279-84

16. Brown EJ, Joiner KA, Garther TA, Hammer CH, Frank MM. 1983. The interaction of C3b bound tp pneumococci with factor H (beta 1H globulin), factor I (C3b/C4b inactivator), and properdin factor B of the human complement system. J. Immunol. 131:409-15

17. Brzoska P, Signer ER. 1991. lpsZ, a lipopolysaccharide gene involved in the symbiosis of Rhizobium meliloti. J. Bacteriol. 173:3235-37

18. Bugert P, Geider K. 1995. Molecular analysis of the ams operon required for exopolysaccharide synthesis of Erwinia amylovora. Mol. Microbiol. 15(5):917-33

19. Caimano MJ, Dillard JP, Yother J. 1995. Molecular evolution of the type 3 capsule locus of Streptococcus pneumoniae. J. Bacteriol. In press

20. Cieslewicz M, Steenbergen S, Vimr E. 1993. Cloning, sequencing, expression and complementation analysis of the Escherichia coli K1 kps region 1 gene, kpsE, and identification of an upstream open reading frame with homology to GutQ. J. Bacteriol. 175:8018-23

21. Costerton JW, Cheng K-J, Geesey GG, Ladd TI, Nickel JC, et al. 1987. Bacterial biofilms in nature and disease. Annu. Rev. Microbiol. 41:435-64

22. Costerton JW, Irvin RT, Cheng K-J. 1981. The bacterial glycocalyx in nature and disease. Annu. Rev. Microbiol. 1981. 35:299-324

23. Cross A. 1990. The biological significance of bacterial encapsulation. Curr. Top. Microbiol. Immunol. 150:87-95

24. Denny TP, Baek SR. 1991. Genetic evidence that extracellular polysaccharide is a virulence factor of Pseudomonas solanacearum. Mol. Plant-Microbe Interact. 4:198-206

25. Dillard JP, Vandersea MW, Yother J. 1995. Characterization of the cassett containing genes for Type 3 capsular polysaccharide biosynthesis in Streptococcus pneumoniae. J. Exp. Med. 181:973-83

26. Dillard JP, Yother J. 1994. Genetic and molecular characterisation of capsule saccharide biosynthesis in Streptococcus pneumoniae type 3. Mol. Microbiol. 12:959-72

27. Dodgson C, Amor P, Whitfield C. 1995. Distribution of the rol gene encoding the regulator of lipopolysaccharide O-chain length in Escherichia coli and its influence on the ssion of group I capsular K antigens. J. Bacteriol. In press

28. Doige CA, Doige G, Ames F-L. 1993. ATP-dependent transport systems in bacteria and humans: relevance to cystic fibrosis and multidrug resistance. Annu. Rev. Microbiol. 47:291-319

29. Dolph PJ, Majerczak DR, Coplin DL. 1988. Characterization of a gene cluster for exopolysaccharide biosynthesis and virulence in Erwinia stewartii. J. Bacteriol. 170:865-71

30. Drake C. 1991. Molecular studies of Escherichia coli capsule gene clusters. PhD thesis. Univ. Leicester, UK

31. Drake CR, Roberts IS, Jann B, Jann K, Boulnois G. 1990. Molecular cloning and expression of the genes encoding the Escherichia coli K4 capsular polysaccharide, a fructose-substituted chondroitin. FEMS Microbiol. Lett. 66:227-30

32. Edwards MS, Kasper DL, Jennings, HJ, Baker CJ, Nicholson-Weller A. 1982. Capsular sialic acid **prevents** activation of the alternative complement pathway by type III, group B streptococci. J. Immunol. 128:1278-83

33. Edwards U, Muller A, Hammerschmidt S, Gerardy-Schahn R, Frosch M. 1994. Molecular analysis of the biosynthesis pathway of the a2,8 polysialic acid capsule by Neisseria meningitidis serogroup B. Mol. Microbiol. 14(1):141-49

34. Fath M, Kolter R. 1993. **ABC transporters** : bacterial exporters. Microbiol. Rev. 7:995-1017

35. Finke A, Bronner D, Nikolaev AV, Jann B, Jann K. 1991. Biosynthesis of the Escherichia coli K5 polysaccharide, a representative of group II capsular polysaccharides: polymerization in vitro and characterization of the product. J. Bacteriol. 173:4088-94

36. Finke A, Jann B, Jann K. 1990. CMPKDO-synthetase activity in Escherichia coli expressing capsular polysaccharides. FEMS Microbiol.

Lett. 69:129-34
37. Finne J. 1982. Occurrence of unique polysialosyl carbohydrate units in glycoproteins of developing brain. J. Biol. Chem. 257:11966-70
38. Frank M, Joiner K, Hammer C. 1987. The function of **antibody** and complement in the lysis of bacteria. Rev. Infect. Dis. 9:S527-45
39. Frosch M, Edwards U, Bousset K, Krausse B, Weisgerber C. 1991. Evidence for a common molecular origin of capsular gene loci in gram-negative bacteria expressing group II capsular polysaccharides. Mol. Microbiol. 5:1251-63
40. Frosch M, Muller A. 1993. Phospholipid substitution of capsular polysaccharides and mechanisms of capsule formation in Neisseria meningitidis. Mol. Microbiol. 8:483-93
41. Frosch M, Muller D, Bousset K, Muller A. 1992. Conserved outer membrane protein of Neisseria meningitidis involved in capsule expression. Infect. Immun. 60:798-803
42. Frosch M, Weisgerber C, Meyer T. 1989. Molecular characterization and expression in Escherichia coli of the gene complex encoding the polysaccharide capsule of Neisseria meningitidis group B. Proc. Natl. Acad. Sci. USA 86:1669-73
43. Ganguli S, Zapata G, Wallis T, Reid C, Boulnois G, et al. 1994. Molecular cloning and analysis of the genes for sialic acid synthesis in Neisseria meningitidis group B and purification of the meningococcal CMP-NeuNAc synthetase enzyme. J. Bacteriol. 176:4583-89
44. Garcia E, Garcia P, Lopez R. 1993. Cloning and sequencing of a gene involved in the synthesis of the capsular polysaccharide of Streptococcus pneumoniae type 3. Mol. Gen. Genet. 239:188-95
45. Glucksmann M, Reuber T, Walker G. 1993. Genes needed for the modification, lymerisation, export and processing of succinoglycan by Rhizobium meliloti: a model for inoglycan biosynthesis. 175:7045-55
46. Goldman RC, Kohlbrenner WE. 1985. Molecular cloning of the structure gene coding for CTP:CMP-3-dexoy-mannooctulosonate cytidyltransferase from Escherichia coli K-12. 163:256-61
47. Govan J. 1988. Alginate biosynthesis and other unusual characteristics associated with the pathogenesis of Pseudomonas aeruginosa in cystic fibrosis. In Bacterial Infections of the Respiratory and Gastrointestinal Mucosae, ed. E Griffiths, W Donachie, J Stephen, pp. 67-96. Oxford: IRL Press
48. Grados O, Ewing VM. 1970. Antigenic relationships between Escherichia coli and Neisseria meningitidis group B. J. Infect. Dis. 122:100-3
49. Guidolin A, Morona JK, Morona R, Hansman D, Paton JC. 1994. Nucleotide sequence analysis of genes essential for capsular polysaccharide biosynthesis in Streptococcus pneumoniae type 19F. Infect. Immun. 62:5384-96
50. Gygi D, Rahman M, Lai H-C, Carlson R, Guard-Petter J, et al. 1995. A cell-surface polysaccharide that facilitates rapid population migration by differentiated swarm cells Proteus mirabilis. Mol. Microbiol. 17:1167-75
51. Hammerschmidt S, Birkholz C, Zarhinger U, Robertson B, Van Putter J, et al. 1994. Contribution of genes from the capsule gene complex (cps) to lipooligosaccharide biosynthesis and serum resistance in Neisseria meningitidis. Mol. Microbiol. 11:885-96
52. Hammerschmidt S, Hilse R, van Putten JPM, Gerardy-Schahn R, Unkmeir R, et al. 1995. Modulation of cell surface sialic acid expression in Neisseria meningitidis via a transposable genetic element. EMBO J. In press
53. Hashimoto Y, Li N, Yokoyama H, Ezaki T. 1993. Complete nucleotide sequencing and molecular characterization of ViaB region encoding Vi antigen in Salmonella typhi. J. Bacteriol. 175:4456-65
54. Higgins CF. 1992. **ABC transporters** : from microorganisms to man. Annu. Rev. Cell Biol. 8:67-113
55. Horwitz MA, Silverstein SC. 1990. Influence of the Escherichia coli capsule on complement fixation and on phagocytosisl. J. Clin. Invest. 65:82-94
56. Howard CJ, Glynn AA. 1971. The virulence for mice of strains of Escherichia coli related to the effects of K antigens on their resistance to phagocytosis and killing by complement. Immunology 20:767-77
57. Huang J, Schell M. 1995. Molecular characterization of the eps gene cluster of Pseudomonas solanacearum and its transcriptional regulation

at a single promoter. Mol. Microbiol. 16(5):977-89
58. Ielpi L, Couso RO, Dankert MA. 1981. Xanthum gum biosynthesis: Pyruvic acid acetal residues are transferred from phosphoenolpyruvate to the pentasaccharide-P-P-lipid. Biochem. Biophys. Res. Commun. 102:1400-8
59. Jann B, Jann K. 1990. Structure and biosynthesis of the capsular antigens of Escherichia coli. Curr. Top. Microbiol. Immunol. 150:19-42
60. Jann K, Dengler T, Jann B. 1992. Corelipid A on the K40 polysaccharide of Escherichia coli 08:K40:H9, a representative of group I capsular polysaccharides. Zentralbl. Bakteriol. 276:196-204
61. Deleted in proof
62. Jann K, Jann B. 1992. Capsules of Escherichia coli, expression and biologic significance. Can. J. Microbiol. 38:705-10
63. Jayaratne P, Keenleyside WJ, MacLachlan PR, Dodgson C, Whitfield C. 1993. Characterization of rcsB rcsC from Escherichia coli 09:K30:H12 and examination of the role of the rcs regulatory system in expression of group I capsular polysaccharides. J. Bacteriol. 175:5384-94
64. Jenkinson FF. 1994. Adherence and accumulation of oral streptococci. Trends Microbiol. 2:209-12
65. Kim K, Kang J, Cross A. 1986. The role of capsular antigens in serum resistance and in vivo virulence of Escherichia coli. FEMS Microbiol. Lett. 35:275-78
66. Kolenbrander PE. 1993. Coaggregation of human oral bacteria: potential role in the accretion of plaque. J. Appl. Bacteriol. 22(5):79-86
67. Kroll S, Hopkins I, Moxon E. 1988. Capsule loss in H. influenzae type b occurs by recombination-mediated disruption of a gene essential for capsular polysaccharide export. Cell 3:347- 56
68. Kroll S, Loynds B, Brophy L, Moxon E. 1990. The bex locus in encapsulated Haemophilus influenzae: a chromosomal region involved in capsular polysaccharide export. Mol. Microbiol. 4:1853-62
69. Kroll S, Loynds B, Moxon E. 1991. The Haemophilus influenzae capsulation gene cluster: a compound transposon. Mol. Microbiol. 5:1549-60
70. Kroll S, Zamze S, Loynds B, Moxon E. 1989. Common organization of chromosomal loci for production of different capsular polysaccharides in Haemophilus influenzae. J. Bacteriol. 171:3343-47
71. Kroncke K, Boulnois G, Roberts IS, Bitter-Suermann D, Golecki J, et al. 1990. Expression of the Escherichia coli K5 capsular antigen: immunoelectron microscopic and biochemical studies with recombinant E. coli. J. Bacteriol. 172:1085-91
72. Laasko D, Homonylo M, Wilmot S, Whitfield C. 1988. Transfer and expression of the genetic determinants for 0 and K antigen synthesis in Escherichia coli 09:K(A)30: and Klebsiella sp. OI:K20 in E. coli K-12. Can. J. Microbiol. 34:987-92
73. Lee CY. 1992. Cloning of genes affecting capsule expression in Staphylococcus aureus strain M. Mol. Microbiol. 6:1515-22
74. Lee CY. 1996. Association of staphylococcal type-1 capsule-encoding genes with a discrete genetic element. Gene. In press
75. Lee JC, Xu S, Albus A, Livolsi PJ. 1994. Genetic analysis of Type 5 capsular polysaccharide expression by Staphylococcus aureus. J. Bacteriol. 176:4883-89
76. Lindahl U, Lidholt K, Spillmann D, Kjellen L. 1994. More to heparin than anticoagulation. Thromb. Res. 75:1-32
77. MacLachlan P, Keenleyside D, Dodgson C, Whitfield C. 1993. Formation of the K30 (group 1) capsule of Escherichia coli 09:K30 does not require attachment to lipopolysaccharide lipid A-core. J. Bacteriol. 175:7515-22
78. Masson L, Holbein BE. 1983. Physiology of sialic acid capsular polysaccharide synthesis in serogroup B Neisseria meningitidis. J. Bacteriol. 154:728-36
79. Masson L, Holbein BE. 1985. Role of lipid intermediate(s) in the synthesis of serogroup B Neisseria meningitidis capsular polysaccharide. J. Bacteriol. 161:861-67
80. Michalek M, Mold C, Bremer E. 1988. Inhibition of the alternative pathway of human complement by structural analogues of sialic acid. J. Immunol. 140:1588-94
81. Morona R, van den Bosch L, Manning PA. 1995. Molecular, genetic and topological characterization of 0-antigen chain length regulation in Shigella flexneri. J. Bacteriol. 177:1059-68

82. Moxon E, Kroll J. 1990. The role of bacterial polysaccharide capsules as virulence factors. Curr. Top. Microbiol. Immunol. 150:65-85

83. Muller P, Keller M, Weng M, Quandt J, Arnold W, et al. 1993. Genetic analysis of Rhizobium meliloti exo YFQ operon: ExoY is homologous to sugar transferases and ExoQ represents a transmembrane protein. Mol. Plant-Microbe Interact. 6:55-65

84. Noel K. 1992. Rhizobial polysaccharides required for symbiosis with legumes. In Molecular Signals in Plant-Microbe Interaction, ed. D Vemia, pp. 341-57. Boca Raton, FL: CRC

85. Ophir T, Gutnick D. 1994. A role for exopolysaccharides in the protection of microorganisms from desiccation. App. Environ. Microbiol. 60:740-45

86. Orskov F. 1984. The genus Escherichia. In Bergey's Manual of Systematic Bacteriology, ed. N Krieg, J Holt, pp. 420-23. Baltimore: Williams & Wilkins

87. Orskov F, Orskov I. 1992. Escherichia coli serotyping and disease in man and animals. Can. J. Microbiol. 38:699-704

88. Orskov F, Sharina V, Orskov I. 1984. Influence of growth temperature on the development of Escherichia coli polysaccharide K antigens. J. Can. Microbiol. 130:2681-84

89. Orskov I, Nyman K. 1974. Genetic mapping of the determinants of two polysaccharide K antigens, K10 and K54, in Escherichia coli. J. Bacteriol. 120:43-51

90. Orskov I, Sharina V, Orskov F. 1976. Genetic mapping of the K1 and K4 antigens (L) of Escherichia coli. Acta Pathol. Microbiol. Scand. Sect. B 84:125-31

91. Pavelka M, Hayes S, Silver R. 1994. Characterization of KpsT, the ATP-binding component of the **ABC - transporter** involved with the export of capsular polysialic acid in Escherichia coli K1. J. Biol. Chem. 269:20149-58

92. Pavelka M, Wright L, Silver R. 1991. Identification of two genes, kpsM and kpsT, in region 3 of the polysialic acid gene cluster of Escherichia coli K1. J. Bacteriol. 173:4603-10

93. Pazzani C, Rosenow C, Boulnois G, Brenner D, Jann K, et al. 1993. Molecular analysis of region I of the Escherichia coli K5 capsule gene cluster: a region involved in cell surface expression of capsular polysaccharide. J. Bacteriol. 175:5978-83

94. Pazzani C, Smith A, Bronner D, Jann K, Boulnois G. 1993. The export of capsular polysaccharides by Escherichia coli. In Polygialic Acid: from Microbes to Man, ed. J Roth, U Rutishauser, F Troy II, pp. 113-23. Basel: Birkhauser Verlag

95. Pearce BJ, Naughton AM, Massure HR. 1994. Peptide permeases modulate transformation in Streptococcus pneumoniae. Mol. Microbiol. 12:881-92

96. Pearce R, Roberts IS. 1995. Cloning and analysis of the gene clusters for the production of the Escherichia coli K10 and K54 antigens: identification of a new group of serA-linked capsule gene clusters. J. Bacteriol. 177:3992-97

97. Petit C, Rigg G, Pazzani C, Smith A, Sieberth V, et al. 1995. Region 2 of the Escherichia coli K5 capsule gene cluster encoding proteins for the biosynthesis of the K5 polysaclcharide. Mol. Microbiol. 17:611-20

98. Pickard D, Roberts M, Maskell D, Hone D, Levine M, et al. 1994. Characterization of defined ompR mutants of Salmonella typhi: OmpR is involved in the regulation of Vi polysaccharide expression. Infect. Immun. 62:3984-93

99. Pigeon RP, Silver RP. 1994. Topological and mutational analysis of KpsM, the hydrophobic component of the **ABC - transporter** involved in the export of polysialic acid in Escherichia coli K1. Mol. Microbiol. 14(5):871-81

100. Raetz CRH. 1990. Biochemistry of endotoxins. Annu. Rev. Biochem. 59:129-70

101. Robbins JB, McCracken GH, Gotschlich EC, Orskov I, Hanson LA. 1974. Escherichia coli K1 capsular polysaccharide associated with neonatal meningitidis. N. Engl. J. Med. 90:267-71

102. Roberson E, Firestone M. 1992. Relationship between desiccation and exopolysaccharide production in soil Pseudomonas sp. Appl. Environ. Microbiol. 58:1284-91

103. Roberts I. 1995. Bacterial polysaccharides in sickness and health. Microbiology 141:2023-31

104. Roberts I, Mountford R, High N, Bitter-Suermann D, Jann K, et al. 1986. Molecular cloning and analysis of the genes for the production of the K5, K7, K12 and K92 capsular polysaccharides in Escherichia coli. J. Bacteriol. 168:1228-33

105. Roberts IS, Saunders FK, Boulnois GJ. 1989. Bacterial capsules and interactions with complement and phagocytes. Biochem. Soc. Trans. 17:462-64

106. Roberts IS, Mountford R, Hodge R, Jann KB, Boulnois GJ. 1988. Common organization of gene clusters for production of differemt capsular polysaccharides (K-antigens) in Escherichia coli. J. Bacteriol. 170:1305-10

107. Roberts M, Roberts I, Korhonen TK, Jann K, Bitter-Suermann D, et al. 1988. DNA probes for K antigen (capsule) typing of Escherichia coli. J. Clin. Microbiol. 26:385-87

108. Rodriguez ML, Jann B, Jann K. 1989. Structure and serological properties of the capsular K4 antigen of Escherichia coli 05:K4:H4, a fructose containing polysaccharide with a chondroitin backbone. Eur. J. Biochem. 177:117-24

109. Rosenow C, Esumeh F, Roberts I, Jann K. 1995. Characterisation and localisation of the KpsE protein of Escherichia coli K5, which is involved in polysaccharide export. J. Bacteriol. 177:1137-43

110. Rosenow C, Roberts I, Jann K. 1995. Isolation from recombinant Escherichia coli and characterization of CMP-KdO synthetase, involved in the expression of the capsular K5 polysaccharide (K-CKS). FEMS Microbiol. Lett. 125:159-64

111. Rubens CE. 1994. Type III capsular polysaccharide of Group B streptococci: role in virulence and the molecular basis of capsule expression. In American Society for Microbiology, ed. VL Miller, B Kaper, DA Portnoy, RR Isberg, pp. 327-39. Washington: Am. Soc. Microbiol.

112. Schmidt A, Jann B, Jann K, Orskov I, Orskov F. 1977. Genetic determinants of the synthesis of the polysaccharide capsular antigen K27(A) of Escherichia coli. J. Gen. Microbiol. 100:355-61

113. Schmidt A, Jann K. 1982. Phospholipid substitution of capsular (K) polysaccharide antigens Escherichia coli causing extraintestinal infections. FEMS Microbiol. Lett. 14:69-74

114. Sieberth V, Rigg G, Roberts I, Jann K. 1995. Expression and characterisation of UDPGLc dehydrogenase (KfiD) which is encoded in the type-specific region 2 of the Escherichia coli K5 capsule genes. J. Bacteriol. 177:4562-65

115. Silver RP. 1994. Genetic analysis of the Escherichia coli K1 capsule gene cluster. In Molecular Genetics of Bacterial Pathogenesis, ed. V. Miller, pp. 313-26. Washington, D.C.: Am. Soc. Microbiol.

116. Silver RP, Aaronson W, Vann WF. 1984. Genetic and molecular analysis of Escherichia coli K1 antigen genes. J. Bacteriol. 157:568-75

117. Silver RP, Annunziato P, Pavelka M, Pigeon R, Wright L, et al. 1993. Genetic and molecular analyses of the polysialic acid gene cluster of Escherichia coli K1. In Polysialic Acid: from Microbes to Man, ed. J Roth, U Rutishauser, F Troy II, pp. 59-72. Basel: Birkhauser Verlag

118. Silver RP, Finn CW, Vann WF, Aaronson W, Schneerson R, et al. 1981. Molecular cloning of the K1 capsular polysaccharide genes of E. coli. Nature 289:696-98

119. Smith AN, Boulnois GJ, Roberts IS. 1990. Molecular analysis of the Escherichia coli K5 kps locus: identification and characterization of an inner-membrane capsular polysacchari transport system. Mol. Microbiol. 4:1863-69

120. Steenbergen SM, Vimr ER. 1990. Mechanism of polysialic acid chain elongation in Escherichia coli K1. Mol. Microbiol. 4:603-11

121. Steenbergen SM, Wrona TJ, Vimr ER. 1992. Functional analysis of sialyltransferase complexes in Escherichia coli K1 and K92. J. Bacteriol. 174:1099-1108

122. Stevens P, Huang SNH, Welch WD, Young LS. 1978. Restricted complement activation by Escherichia coli with the K1 capsular serotype: a possible role in pathogenicity. J. Immunol. 121:2171-80

123. Stevenson G, Lee S-J, Romana LK, Reeves PR. 1991. The cps gene cluster of Salmonella strain LT2 includes a second mannose pathway: sequence of two genes and relationship to genes in the rjb gene cluster. Mol. Gen. Genet. 227:173-80

124. Strominger JL. 1973. The actions of penicillin and other antibiotics on bacterial cell wall synthesis. Johns Hopkins Med. J.

133:63-81
125. Deleted in proof
126. Troy FA, Cho J-W, Ye J. 1993. Polysialic acid capsule synthesis and chain translocation in neuroinvasive Escherichia coli K1: "activated" intermediates and a possible bifunctional role for the undercaprenol. In Polysialic Acid: from Microbes to Man, ed. J Roth, U Rutishauser, F Troy II, pp. 93-111. Basel: Birkhauser Verlag
127. Troy FA, Frerman FE, Heath EC. 1971. The biosynthesis of capsular polysaccharide in Aerobacter aerogenes. J. Biol. Chem. 246:118-33
128. Van Eldere J, Brophy L, Loynds, B, Celis P, Hancock I, et al. 1995. Region II of the Haemophilus influenzae Type b capsulation locus involved in serotype-specific polysaccharide synthesis. Mol. Microbiol. 15(1):107-18
129. Vann WF, Jann K. 1979. Structure and serological specificity of the K13 antigenic polysaccharide (K13 antigen) of urinary tract infective E. coli. Infect. Immun. 25:85-92
130. Vann WF, Schmidt MA, Jann B, Jann K. 1981. The structure of the capsular polysaccharide (K5 antigen) of urinary tract infective Escherichia coli O10:K5:H4. A polymer similar to desulfo-heparin. Eur. J. Biochem. 116:359-64
131. Vann WF, Soderstrom T, Egan W, Tsui F-P, Schneerson R, et al. 1983. Serological, chemical and structural analyses of the Escherichia coli cross-reactive capsular polysaccharides K13, K20 and K23. Infect. Immun. 39:623-29
132. Vann WF, Zapata G, Roberts I, Boulnois G, Silver R. 1993. Structure and function of enzymes in sialic acid metabolism in polysialic acid producing bacteria. In Polysialic Acid: from Microbes to Man, ed. J Roth, U Rutishauser, F Troy II, pp. 125-36. Basel: Birkhauser Verlag
133. Vimr ER. 1991. Map position and genome organization of the kps cluster for polysialic acid biosynthesis in Escherichia coli K1. J. Bacteriol. 173:1335-38
134. Vimr ER, Aaronson W, Silver RP. 1989. Genetic analysis of chromosomal mutations in the polysialic acid gene cluster of Escherichia coli K1. J. Bacteriol. 171:1106-17
135. Vimr ER, Steenbergen S, Cieslewicz S. 1995. Biosynthesis of the polysialic acid capsule in Escherichia coli K1. J. Ind. Microbiol.
136. Vimr E, Troy FA. 1985. Regulation of sialic acid metabolism in Escherichia coli: role of N-acylneuraminate pyruvatelyase. J. Bacteriol. 164:854-60
137. Ward C, Inzana T. 1995. Molecular analysis of a DNA region involved in the export of the capsular polysaccharide of Actinobacillus pleuropneumoniae serotype 5 (Abstract B232)-95th Annu. Meet. Am. Soc. Microbiol. Washington, D.C.: Am. Soc. Microbiol. Abstr. B232
138. Whitfield C, Schoenhals G, Graham L. 1989. Mutants of Escherichia coli 09:K30 with altered synthesis and expression of the capsular K antigen. J. Gen. Microbiol. 135:2589-99
139. Whitfield C, Troy FA. 1984. Biosynthesis and assembly of the polysialic acid capsule in Escherichia coli K1. J. Biol. Chem. 259:12769-75
140. Whitfield C, Valvano M. 1993. Biosynthesis and expression of cell-surface polysaccharides in gram-negative bacteria. Adv. Microbiol. Phys. 35:135-46
141. Wunder D, Aaronson W, Hayes S, Bliss J, Silver R. 1994. Nucleotide sequence and mutational analysis of the gene encoding KpsD, a periplasmic protein involved in the transport of polysialic acid in Escherichia coli K1. J. Bacteriol. 176:4025-33
142. Wyle F, Artenstein M, Brandt BL, Tramont EC, Kapser DL, et al. 1972. Immunologic response of man to group B meningococcal polysaccharide vaccines. J. Infect. Dis. 126:514-21
143. Yamada M, Yamada Y, Saier M Jr. 1990. Nucleotide sequence and expression of the gutQ gene within the glucitol operon of Escherichia coli. J. DNA Seq. Mapp. 1:141-45
144. Yanofsky C. 1981. Attenuation in the control of expression of bacterial operons. Nature 289:751-58
145. Zapata G, Vann W, Aaronson W, Lewis M, Moos M. 1989. Sequence of the cloned Escherichia coli K1 CMP-N-acetyl neuraminic acid synthetase gene. J. Biol. Chem. 64:14769-74

DESCRIPTORS:

Bacterial capsules; Metabolism--Carbohydrates; Bacterial genetics

03293657 H.W. WILSON RECORD NUMBER: BGSI96043657 (THIS IS THE FULLTEXT)
Mechanisms of adhesion by oral bacteria.
Whittaker, Catherine J
Klier, Christiane M; Kolenbrander, Paul E
Annual Review of Microbiology (Annu Rev Microbiol) v. 50 ('96) p. 513-52
SPECIAL FEATURES: bibl il ISSN: 0066-4227
LANGUAGE: English
COUNTRY OF PUBLICATION: United States
RECORD TYPE: Abstract; Fulltext RECORD STATUS: Corrected or revised record
WORD COUNT: 16881

ABSTRACT: Adherence to a surface is a key element for colonization of the human oral cavity by the more than 500 bacterial taxa recorded from oral samples. Three surfaces are available: teeth, epithelial mucosa, and the nascent surface created as each new bacterial cell binds to existing dental plaque. Oral bacteria exhibit specificity for their respective colonization sites. Such specificity is directed by adhesin-receptor cognate pairs on genetically distinct cells. Colonization is successful when adherent cells grow and metabolically participate in the oral bacterial community. The potential roles of adherence-relevant molecules are discussed in the context of the dynamic nature of the oral econiche. Reprinted by permission of the publisher.

TEXT:

INTRODUCTION

Adherence mechanisms of oral bacteria are essential to bacterial colonization of the oral cavity. In their absence, bacteria become part of the salivary flow and are swallowed. Consequently, oral bacteria have evolved several mechanisms to fulfill this role. The mechanisms are highly specific in that the oral cavity is colonized principally by bacteria that are only found in the oral cavity. For example, bacteria of the intestinal flora are found in the oral cavity in very low numbers, if at all. Likewise, oral bacteria are found infrequently in other ecosystems. They can, however, cause infections in other regions of the body, and the oral cavity is often considered a reservoir for these infectious bacteria. Although humans have ample opportunity for exchange of bacteria among us, we tend to maintain our own personal flora even after antibiotic **therapy** (153).

Three principle surfaces are available for oral bacterial colonization. Both teeth and mucosal epithelial cells are coated with bacteria. The third surface is the nascent bacterial layer that constitutes dental plaque on teeth and the initial coating on epithelial cells. Indeed, bacteria recognize and bind to other kinds of bacteria in suspension and on substrata. Accretion of bacteria onto a surface forms a nascent layer, while adherence of bacteria in suspension forms mixed cell--type coaggregates. Coaggregation was first reported by Gibbons & Nygaard in 1970 (66) and is defined as the recognition and adhesion between genetically distinct bacteria (105). Suspended coaggregates may also accrete and contribute to the formation of dental plaque (190).

This review includes a discussion of the developments in coaggregation; it was first reviewed in 1988 (102). Adhesins and complementary receptors are described, and other forms of oral bacterial adherence such as binding to teeth and epithelial cells are coordinated within the framework of an oral microbiological econiche. Finally, a newly developing area of invasion of host cells by oral bacteria is reviewed.

Several other reviews of adherence of oral bacteria should be consulted for a complete discussion of this topic. The field of oral bacterial adherence is very active and has expanded considerably in the last seven years. For additional discussions on specific topics, the reader is referred to some excellent reviews on oral bacterial invasion (57), streptococcal receptors and molecular modeling (21), streptococcal adherence mechanisms (76, 87, 88), Streptococcus mutans group (16), ecology of coaggregations (118, 119), surface structures of oral bacteria (75), and saliva-bacterium interactions (181).

SURFACES, OCCUPANTS, AND MULTIPLE MECHANISMS OF ADHERENCE

More than 500 bacterial taxa have been recorded in samples taken from the

human oral cavity, and about 300 have been named (153). The most prevalent oral bacteria are members of the 22 genera listed in Table 1. Although 13 genera are gram-negative, a single gram-positive (Actinomyces naeslundii) and a gram-negative (Fusobacterium nucleatum) species predominate in dental plaque under nearly all conditions of health or disease (153). Surveys have been conducted that encompassed at least 1000 strains representing these genera. Strains from each genus coaggregate with a specific set of partner strains, usually from a different genus. Mixing a pair of partners results in clumps or coaggregates composed of an interacting network of the two cell types. By contrast, simple cellular agglutination or aggregation refers to interactions between cells that are genetically identical. Most and probably all oral bacteria coaggregate with at least one partner cell type. The first report, or in some cases the first extensive report, of coaggregations involving members of each genus is listed. Numerous additional studies have been conducted for most of the genera, and those reports are discussed in the various sections of this review.

While coaggregations have been studied since 1970, invasion of epithelial cells by oral bacteria has been examined only recently and was first reported in the early 1990s for suspected periodontal pathogens Actinobacillus actinomycetemcomitans (152) and Porphyromonas gingivalis (127). A third potential periodontopathogen, Treponema denticola, may not actually invade cells but rather insert its chymotrypsin-like proteinase (205).

Each oral bacterium has the capability of several mechanisms for adherence; two examples follow. P. gingivalis strains possess the armament to coaggregate (112), bind to saliva-coated hydroxyapatite (SHA) (27), hemagglutinate (24), and adhere to and invade epithelial cells (123). This bacterium also binds to several matrix molecules, like fibronectin (130), fibrinogen (131), and collagen (159), and produces proteases that may promote adherence (5). Streptococcus gordonii PK488 adheres to SHA and coaggregates with A. naeslundii PK606 (10, 109, 110), other streptococci (113), and fusobacteria (112) by different mechanisms. Mutants of strain PK488 that fail to coaggregate with PK606 retain the lactose-inhibitable coaggregations with streptococci and the lactose-noninhibitable coaggregations with fusobacteria, and they bind to SHA. Thus, multiple adhesins on a given cell are likely to mediate distinct interactions with different surfaces, which can be animate or inanimate.

POPULATION COMPOSITION CHANGES

Left undisturbed, our teeth accrete a thick layer of plaque, a coating of bacteria mixed with salivary and serum molecules. Immediately following professional teeth cleaning, a thin host-derived layer called the acquired pellicle covers the tooth surface. The pellicle consists of numerous components, including mucins, glycoproteins, proline-rich proteins, histidine-rich proteins, enzymes like alpha-amylase, phosphate-containing proteins like statherin, and other molecules. The acquired pellicle coating appears to mask the contribution of the substratum surface, for amalgam, hydroxyapatite, and titanium were colonized in vivo by the same kinds of bacteria (134). Bacteria rapidly accrete and in turn become an available surface for adherence by other bacteria. Bacteria are immersed in salivary and serum molecules that they can utilize as nutrients (39, 202). This combination of recognition, accretion, and growth leads to successful colonization (16a, 190, 191).

TEMPORAL RELATIONSHIP OF SUBGINGIVAL OCCUPANTS AND COAGGREGATION PARTNERSHIPS

An organism may be able to adhere but unable to grow in a particular lotic or flowing environment. When the environment becomes more favorable, the organism may proliferate and constitute a major portion of the biofilm. Consider a possible scene of repopulation of a tooth surface after cleaning. The acquired pellicle covers the tooth. Streptococci, the principle early colonizers, bind to acidic proline-rich-proteins (65, 81) and other receptors like alpha-amylase (182, 183) and sialic acid (44, 81) in the acquired pellicle. Streptococci also participate in intrageneric coaggregation (81, 113, 124), which offers an extra advantage in allowing them to bind to the nascent monolayer of already bound streptococci (165, 191). In addition, actinomyces, which are other primary colonizers, bind

to the acquired pellicle (64) and to the streptococci (see 102, 103, 115, 118).

Each streptococcal and actinomyces strain binds specific salivary molecules. Thus, from a common pool of salivary molecules, each strain of early colonizer may be coated with distinct molecules. Identical cells coated with a specific salivary molecule may agglutinate, which would lead to a microconcentration and juxtapositioning of a particular strain. Alternatively, growth of a particular accreted strain would lead to a microcolony coated with specific salivary molecules. Such events could dramatically alter the diversity of salivary molecules exposed to later colonizers. In this way, early colonizers may dictate small adjustments in the temporal accretion of later colonizers.

Both streptococci and actinomyces are facultatively anaerobic, and doubling times for microbial populations during the first four hours of plaque development are less than one hour (209). Consequently, these two groups of primary colonizers are thought to prepare the environment for later colonizers that have more fastidious requirements for growth and that grow more slowly (209). Other bacteria like fusobacteria, veillonellae, propionibacteria, rothiae, capnocytophagae, and prevotellae bind to streptococci and/or actinomyces (see 118). Each new accreted cell becomes a nascent surface and therefore becomes a coaggregation bridge to the next potentially accreting cell type that passes by. The accreted cells metabolize and the econiche changes. Cells that could not survive the originally aerobic environment before are now capable of colonizing the anaerobic plaque, but they require attachment points. We have proposed that a principal one is the anaerobic fusobacteria, which as a group coaggregates with all oral bacteria. Accordingly, the fusobacteria act as bridges to anchor the environmentally more-fastidious late colonizers.

The evidence for in vivo implementation is based on the fact that early colonizers coaggregate among each other and with fusobacteria. Late colonizers coaggregate with fusobacteria but do so infrequently among themselves (see 119 for further discussion). Some interactions between late colonizers have been reported (71, 218). In a study of Sudanese adult periodontitis patients, Prevotella intermedia was never found unless F. nucleatum was also present (6). Late colonizers like P. gingivalis and T. denticola are found in close association in vivo (96, 188). Nutritional communication with metabolic end products has been reported between T. denticola and P. gingivalis (72). Considering the kinds of bacteria successful in initial adherence to the acquired pellicle, the highly specific coaggregation partnerships, and the temporal colonization profiles of bacteria reveal that all are essential for participation in development of stable dental plaque. Other factors are also essential for successful colonization. For example, metabolic communication among the community and evasion of host defenses play active roles.

Fusobacteria are not numerous in the initial stages of colonization of the tooth surface, but they increase to about 50[percent] of the numbers of either actinomyces or streptococci in normal plaque. They can bind to statherin (64), a component of the acquired pellicle, but their ability to coaggregate with members of all the genera makes them unusual (112). Each Fusobacterium nucleatum strain has its own set of partners, which includes intrageneric coaggregation with other fusobacteria (120). Considering fusobacteria in total, they coaggregate with all other human oral bacteria so far tested (97, 98, 108, 112, 120). Many of these coaggregations are lactose-inhibitable (97, 108, 120), which adds to the already extensive number of lactose-inhibitable coaggregations among other early colonizers (118). The thought of using a rinse of lactose to disrupt dental plaque should be dismissed, because there are an equal or greater number of coaggregations that are insensitive to lactose. Secondly, lactose, like sucrose, is a metabolizable sugar that leads to plaque growth and acid production rather than inhibition of plaque formation. The diversity of oral bacteria and their multiple adherence mechanisms leads one to think of dental plaque as a dynamic biofilm. Daily oral hygiene disrupts the bacterial community, which reforms daily with readily available nutrients both from food intake and host-derived salivary and serous molecules.

SALIVARY PROTEINS/PELLICLE RECEPTORS

The bacteria that are present in the oral cavity are in constant contact with saliva. As soon as an organism enters the mouth, it becomes coated

with specific salivary proteins, which can increase the adhesion of the bacteria to oral tissues. As well as providing receptors for adhesion, saliva can aggregate bacteria (41) and may aid in their removal. Oral bacteria recognize different receptors within the salivary pellicle (41). Many experiments are done by coating hydroxyapatite, a model for the tooth surface with saliva or specific saliva components.

Actinomyces, porphyromonads, and streptococci bind to acidic proline-rich proteins (PRP's) attached to hydroxyapatite. In in vitro assays, the type 1 fimbriae expressed by Actinomyces viscosus interact with several proline-rich salivary molecules coating either hydroxyapatite or polystyrene (31). Immobilized proline-rich glycoproteins (PRG) have been shown to promote adhesion of some oral bacteria to hydroxyapatite, e.g. A. viscosus (31), F. nucleatum (67), and S. gordonii (136). Multiple salivary components, including the low molecular-weight salivary mucin and highly glycosylated proline-rich glycoproteins (156), a-amylase (156, 183) and proline-rich peptides (65), promote adhesion of streptococci.

Several species of oral streptococci including S. gordonii bind salivary a-amylase (183), a plaque constituent. Only animals exhibiting salivary amylase activity in their saliva harbor amylase-binding streptococci. Amylase-binding bacteria are numerous in supragingival plaque, suggesting that this enzyme may serve as a pellicle receptor.

Fimbriae are composed of structural subunits called fimbrillin and are required for the attachment of P. gingivalis to whole saliva-coated hydroxyapatite beads (133). P. gingivalis fimbriae are important in virulence, for they possibly mediate initial adherence and colonization of the oral cavity. Recombinant P. gingivalis fimbrillin protein (r-Fim) was able to bind to SHA (186). PRP1 and statherin were found to significantly enhance the binding of r-Fim to hydroxyapatite. The C-terminal region of the fimbrillin subunit of fimbriae appears to be responsible for this binding. These results suggest that P. gingivalis fimbriae bind strongly through protein-protein interactions to PRPs and statherin molecules bound to surfaces.

Different streptococcal species preferentially colonize different oral sites and coadhere to a varied range of bacteria, so it is possible that each of the various oral streptococci have evolved unique adhesins. Hence, there appear to be homologous adhesins present in streptococci, but their functions have been honed so that they cooperate, not compete, with each other for binding sites. Mucin-like salivary glycoprotein interacts with many species of oral streptococi and may function in regulating streptococcal colonization of oral tissues. This interaction may induce a nonimmunologic mechanism for clearing these organisms from the oral cavity. Alternatively, salivary agglutinin that is bound to oral tissue surfaces may promote streptococcal adherence to these tissues (18, 101). Among the oral streptococci, much research is conducted on the mutansgroup streptococci because of their cariogenic potential. Streptococcus mutans secrete glucosyltransferase, which enzymatically converts sucrose to glucans, and both the enzyme and glucan are integral parts of the acquired pellicle. Adsorbing glucosyltransferase to SHA and incubating in the presence of sucrose promotes the formation of glucans, which can function as specific binding sites for S. mutans (187).

STREPTOCOCCAL ADHERENCE INTERACTIONS

During the transition from colonization predominantly by streptococci and actinomyces in the first few hours to later colonizing genera, a vast array of surface molecules are presented to the environment. As each new cell type adheres, its cell body becomes a nascent surface presented to the environment. In this way, dental plaque quickly presents numerous possible receptors and adhesins available for specific recognition among different strains. Such interactions are presumed to occur by some of the same molecules as those that mediate coaggregations. This idea is shown diagrammatically in Figure 1. The four topics discussed are intrageneric coaggregation, intergeneric coaggregation, multigeneric coaggregation, and accretion.

Figure 1 is not intended to be comprehensive but rather illustrative of possible interactions of streptococci with their changing environment. The mediation of interactions represented by identical symbols is the

simplest interpretation of the available data. Several streptococci are named and shown, but the focus of this discussion is on S. gordonii PK488, which coaggregates with A. naeslundii PK606 (109). This intergeneric coaggregation is mediated by a putative adhesin (obelisk with stem) on the streptococcus that recognizes a complementary receptor (obelisk) on the actinomyces.

In Figure 1, several types of coaggregations are shown to be mediated by the same receptor (the solid black rectangle). For example, intrageneric coaggregation between S. gordonii PK488 and either S. oralis 34 or Streptococcus SM PK509 is depicted as mediated by these surface components. COG- mutants of S. gordonii PK488, selected by their failure to coaggregate with one of the streptococci (represented by solid black rectangle), lose their ability to coaggregate with the other streptococcus but retain it with A. naeslundii PK606 (represented[cont. on p.522] by obelisk). Likewise, S. gordonii DL1 and Streptococcus sanguis 12 (both are depicted in Figure 1 directly above S. gordonii PK488) also coaggregate with both Streptococcus SM PK509 and S. oralis 34 in indistinguishable lactoseinhibitable coaggregations. Although no mutant of Streptococcus SM PK509 has been isolated, it is expected that one selected by its failure to coaggregate with S. gordonii PK488 [altered in the receptor (solid black rectangle)] would also be unable to coaggregate with either S. gordonii DL1 or S. sanguis 12. Such COG- mutants of S. oralis 34 have been isolated, and they predictably lost their ability to coaggregate with S. gordonii DL1 and S. sanguis 12, but retained coaggregation with Streptococcus SM PK509.

S. gordonii DL1 COG- mutants selected on the basis of inability to coaggregate with S. oralis 34, which is a streptococcal strain similar to S.oralis C104, are also unable to coaggregate with S. oralis C104 and Streptococcus SM PK509, but they retain the ability to coaggregate with A. naeslundii PK606 (33). Several spontaneous mutants were tested, and each failed to express a 100-kDa surface protein (32). S. gordonii PK488 also expresses a 100-kDa surface protein that reacts with an antiserum that recognizes the 100-kDa protein on S. gordonii DL1 (32). This antiserum blocks coaggregation between S. gordonii DL1 and S. oralis 34, S. oralis C104, or Streptococcus SM PK509, but it has no effect on the intergeneric coaggregation between S. gordonii DL1 and A. naeslundii PK606 (32). These results suggest that many intrageneric interactions may be functionally similar and are mediated by structurally similar molecules on the coaggregation partners.

The intergeneric coaggregations between S. gordonii DL1 and either A. naeslundii PK606 (Figure 1, upper right) or Propionibacterium acnes PK93 (Figure 1, lower right) are also shown to be mediated by the (solid black rectangles). This feature distinguishes the adhesins (rectangles with stem) on the three streptococci. Of these three streptococci, only S. gordonii DL1 exhibits lactose-inhibitable coaggregation with A. naeslundii PK606 and P. acnes PK93. Thus, the lactose-sensitive adhesin on S. gordonii DL1 has a broader specificity than the adhesin on the other two streptococci. Spontaneous COG- mutants of S. gordonii DL1 that were selected on the basis of failing to coaggregate with streptococci also lost the lactose-inhibitable coaggregations with A. naeslundii PK606 and P. acnes PK93 (33). These data suggest that there is only one putative lactose-sensitive adhesin on S. gordonii DL1. Although the receptor (solid black rectangle) on S. oralis 34 and Streptococcus SM PK509 may be recognized by all three streptococci, (S. sanguis 12, S. gordonii DL1, and S. gordonii PK488), the receptors (solid black rectangles) on the other partners (A. naeslundii PK606, and P. acnes PK93) are not likely to be identical to the homologous S. oralis 34 and Streptococcus SM PK509 receptors.

The fourth mode of adherence shown in Figure 1 is accretion to the tooth surface (lower left). Four streptococcal strains are shown, and all express an adhesin (semicircle with stem) that recognizes a receptor (shaded ellipse). All strains are known to bind to saliva-coated hydroxyapatite (26, 53, 59, 60, 110, 167). It is not known whether all four bind to the same or different receptors of the acquired pellicle, but all four express the putative adhesin ScaA or a homolog (10, 110). The putative adhesin (semicircle with stem) is discussed in detail below. The molecular nature of the receptor(s) is unknown. Streptococcus parasanguis FW213 does not coaggregate with other streptococci or with A. naeslundii PK606 (110), so it is depicted in Figure 1 with a single adhesin; whereas the other three streptococci are shown with three kinds, each recognizing a

distinct receptor. Several sophisticated chemical analyses of streptococcal receptors have been reported recently (1-3, 20, 69, 170, 171) and are discussed below. Colonization of S. gordonii DL1 (Challis) in the oral cavity of mice occurred in a lactobacillus-free mouse but not in a conventional mouse (139).

ADHESINS

Most known oral bacterial adhesins are from streptococci (Table 2). The subunit size ranges from 35 to 380 kDa. The adhesins are presented at the surface by at least two distinct mechanisms. The first includes many proteins in the range of 210-260 kDa that contain the sequence LPxTG near the C terminus, which may be important for proper sorting in the cell wall (185). A model incorporating the peptide cleavage between the T and G of the LPxTG motif, in which the cleavage has similarity to the transpeptidation reaction of cell wall synthesis, has been proposed (161). This model proposes the mechanisms whereby the C-terminal T can be linked to the following: the N-terminal amino acid of the peptide cross-bridge, the free amino group of the diaminopimelic acid, or lysine of the tetrapeptide attached to the carbohydrate backbone of the bacterial cell wall.

The second mechanism of surface presentation is based on the fact that many of the proteins of 35-78 kDa contain the sequence LxxC, which is the consensus motif for lipoproteins (77, 78). This motif is located about 20 amino acids from the N-terminus of the prolipoprotein and is the site of cleavage by signal peptidase II, an integral membrane protein that catalyzes the cleavage and yields cysteine as the N-terminal amino acid. The cysteine become lipid-modified, and surface exposure of these lipoproteins appears to occur by anchoring the amino terminus in the cytoplasmic membrane and by exposure of the C-terminal region. Thus, there appear to be two distinct presentations: (a) peptidoglycan linkage and surface exposure of the N-terminal region of the adhesin, and (b) cytoplasmic membrane anchoring and surface exposure of the C-terminal region of the adhesin. [cont. on p.527]

The streptococcal adhesins CshA and CshB bind to actinomyces cell surface molecules (148, 149). They are antigenically related, are produced by S. gordonii DL1 (Challis), and are encoded by genes at distinct chromosomal loci (149). CshA has four regions of interest. The precursor form has a 41 amino acid leader peptide, a mature peptide N-terminal segment of residues 42-878, 13 repeating blocks of 101 amino acids and three shorter blocks to comprise amino acids 879-2417, and a C-terminal anchor domain with a sequence LPxTG for proper anchoring to the cell wall (161). In fact, CshA has four additional LPxTG sequence motifs that occur at the same position in their four respective repeating 101 amino acid blocks (149). Insertional mutations in cshA (which encodes CshA) caused reduced cell-surface hydrophobicity and reduced ability to coaggregate with strains of A. naeslundii (148). Mutations in cshB (which encodes CshB) had less effect on hydrophobicity and coaggregation. Both CshA and CshB were required to confer S. gordonii with the ability to colonize the murine oral cavity (149).

The exciting discovery of lipoproteins on the surface of oral streptococci (86, 176) has resulted in an awareness of other lipoproteins among putative adhesins (10, 54, 59, 86, 89, 110, 176, 177). Streptococcal coaggregation adhesin ScaA is a 34.7-kDa lipoprotein (110) expressed by S. gordonii PK488, which coaggregates with A. naeslundii PK606 (109). It is a member of the lipoprotein receptor antigen (Lra) group 1 (88). The genes for six of these proteins have been identified in six species (Table 2). Reactive fragments were found by southern hybridization with a scaA-probe in 12 oral streptococci (10). These streptococci also expressed surface proteins that migrated as 36-38-kDa molecular size in sodium dodecylsulfate-polyacrylamide gel electrophoresis gels, and they were detected in immunoblots using antiserum that blocks coaggregation between S. gordonii PK488 and A. naeslundii PK606 (10). COG- mutants of S. gordonii PK488 that do not coaggregate with A. naeslundii PK606 also do not produce this protein (109).

The other five members of group 1 include FimA, which is a 36-kDa protein associated with the fimbriae of S. parasanguis FW213, which can block attachment of this organism to SHA (167). Both insertion of the aph-3 kanamycin resistance gene cassette into fimA and deletion of fimA of S. parasanguis FW213 reduced binding of the streptococci to fibrin

monolayers and reduced the virulence of the streptococci in a rat model of endocarditis (19). FimA shows 87[percent] homology to a 34.7-kDa saliva binding protein (SsaB) from S. sanguis 12 (60). The scaA sequence shows 86[percent] and 73[percent] homology with the ssaB and fimA adhesin genes respectively (110). A 37-kDa surface protein PsaA has been identified from Streptococcus pneumoniae, which shows 80[percent] homology with SsaB, and 92.3[percent] homology with FimA (177). A 34.7-kDa surface protein EfaA from Enterococcus faecalis also showed between 55-60[percent] homology with this group of streptococcal proteins (142). The nucleotide sequence of the homologous scbA gene of Streptococcus crista, which forms corncob coaggregates with fusobacteria, was reported (35).

No function has, so far, been assigned to scbA, psaA, and efaA. All six of these lipoproteins form a homologous group based on sequence. While the functions of some are proposed, purification of the necessary amount of these proteins has hampered complete characterization of their functions.

Recently, two additional homologous proteins in quite different bacteria have been reported. A manganese transport protein of the cyanobacterium, Synechocystis 6803 is 30[percent] identical (54[percent] similar) in sequence to ScaA (14). The nucleotide sequence of the gene encoding a 31-kDa, rare outer membrane protein from Treponema pallidum subsp. pallidum predicts repeated stretches of amphipathic beta-sheets typical of membrane-spanning sequences of outer membrane proteins (15). The deduced Tromp 1 protein sequence is 29[percent] identical and 53[percent] similar to ScaA. Evidence for homologs in such varied bacteria suggests that these proteins may have evolved distinct functions needed by their respective organisms. On the other hand, these proteins may be essential for a common function such as maintaining surface layers in proper orientation, and their loss, if not fatal, may have a causal effect on adherence.

The DNA sequence of the clone containing scaA indicates that scaA is part of an ATP binding--protein cassette (ABC) system found originally to constitute the binding-protein--dependent transport systems of gram-negative enteric bacteria (9, 80, 84, 164) and the lipoprotein transporters of oligopeptides and sugars of gram-positive bacteria (7, 8, 176). The **ABC** system of **transporters** is found in bacteria and eukaryotes (79) and consists of three basic parts: one or two ATPases, one or two hydrophobic membrane proteins, and one substratespecific binding protein. In gram-negative bacteria the binding proteins are found in the periplasm, but in gram-positive bacteria the binding lipoproteins were proposed to be anchored in the cytoplasmic membrane (68), which has been shown (7). The lipoproteins of gram-positive bacteria have been recently reviewed (197). They are lipid-modified at the N-terminus of the mature protein and are thus anchored to the membrane. This leaves the C-terminal region available for binding to its cognate receptor.

The idea that a streptococcal lipoprotein may also be important in adherence functions was proposed by Jenkinson in 1992 (86) and later expanded by our laboratory to involve an ABC system (110) consisting of ScaA lipoprotein adhesin, hydrophobic membrane protein, and ATP-binding protein encoded by a putative ABC operon (11, 110). Considering that substrate-binding lipoproteins may recognize their cognate ligand either at the cell surface or within the porous peptidoglycan layer, it is not difficult to envision that these lipoproteins may also recognize the identical sequence of the cognate ligand when it is part of a receptor on the surface of a coaggregating partner cell. In addition, calcium ions, which are required for coaggregation, may be bound by a lipoprotein, as occurs for the Mn transporter of Synechocystis 6803 (14).

Binding between the coaggregation mediators has been thought to occur externally to the peptidoglycan at the bacterial cell surfaces. The net negative charge of the bacterial cell surface is often proposed as a barrier to cell-to-cell contact phenomena. However, just as soluble oligopeptides and sugar substrates are able to penetrate the porous peptidoglycan layer, it is proposed that fibrous molecules (capsular and fimbrillar) on a coaggregating partner cell surface can be recognized by a cytoplasmic membrane-anchored lipoprotein adhesin. Even a relatively small molecule like the 34.7-kDa ScaA lipoprotein has 310 amino acids, and an alpha-helix of 21 amino acids is sufficient to span about 30 A (or 3.0 nm) (51). So, the logistics of presenting the C-terminal region to a point external or near the cell surface does not seem insurmountable for a membrane-anchored lipoprotein.

A set of three 76-78-kDa lipoproteins, called HppAGH for hexa- or heptapeptide permease, from S. gordonii DL1 was necessary for cell growth on peptides of 5-7 amino acid residues (89). It was proposed that they form a binding-protein complex for uptake of primarily hexa-heptapeptides. Inactivation of the hppA gene caused reduced growth rate of cells, affected competence onset and transformation efficiency, and caused an increase in aminopterin resistance. These three proteins are highly homologous to the AmiA, AliA (PlpA), and AliB lipoproteins of an oligopeptide permease system in S. pneumoniae (7, 8, 168), which also affect competence and transformation efficiency. Mutations in amiA and plpA (aliA) confer greater than 50[percent] decrease in pneumococcal adherence to epithelial and endothelial cells (37). HppA (originally SarA; see 90) was first reported as a protein involved in serum-induced cell aggregation, saliva-mediated aggregation, and coaggregation with some actinomyces strains.

Coaggregations among streptococci are very common and are often inhibited by GalNAc (113). One of these streptococci, S. gordonii DL1, bears the putative adhesin that recognizes a GalNAc-containing carbohydrate receptor on several streptococci (32, 33). Transposon Tn916 was used to insertionally inactivate coaggregation-relevant genes of S. gordonii DL1. COG- mutants were unable to coaggregate with the streptococcal partners but retained the ability to coaggregate with partners belonging to other genera. The region flanking the Tn916 insertion was sequenced and used to identify a 0.5-kb EcoRI chromosomal fragment, which was cloned into pQ (143), an E. coli-streptococcal insertion vector. Insertion mutants showed altered coaggregation with streptococci, but again retained wild-type coaggregation properties with other genera of bacteria. DL1 expressed a 100-kDa surface protein that was absent in both the Tn916 and pQ insertion mutants (215), as had also been observed with spontaneous COG- mutants (32). This is the first report of a streptococcal adhesin mediating intrageneric coaggregation.

Streptococcus gordonii G9B expresses an 80-kDa antigen complex that is necessary for its adherence to SHA (129, 172). A 153-kDa protein on the surface of S. gordonii G9B that mediates bacterial adhesion to human endothelial and epithelial cells in vitro has also been identified (206). The purified adhesin has glucosyltransferase activity, is able to bind directly to host cells, and is also able to block subsequent adhesion of viable streptococci. Adhesion of S. gordonii to endothelial cells can be inhibited with low molecular weight dextrans and **antibodies** specific to the 153-kDa protein. GTF hydrolyzes sucrose and catalyzes the formation of water soluble and water insoluble glycans, which may mediate bacterial binding to the tooth surface. GTF synthesis is positively regulated by the immediately upstream gene, rgg, in S. gordonii (207). Both rgg and the GTF structural gene, gtfG, are found on the same restriction fragment in strains of S. gordonii, S. sanguis, and S. oralis (207). However, no rgglike determinants were found in mutans streptococci, Streptococcus mitis and Streptococcus salivarius (207). The significance of these important findings to colonization of the oral streptococci remains to be elucidated.

Extensive regions of sequence homology are found among the polypeptides comprising a group of surface proteins referred to as I/II antigens, P1-like polypeptides, and antigen B, which are called the antigen I/II family (Table 2) in this review. They are in the molecular size range of 165-210 kDa and are members comprising a conserved gene family (144) found in Streptococcus sobrinus, S. mutans, and S. gordonii strains. They are often functionally different. For example, in S. mutans KPSK2 the interaction of the polypeptide MSL-1 with salivary agglutinin is inhibited by fucose and lactose (41), but in S. gordonii the interactions of SSP5 (44, 124) and SspA (91) with salivary agglutinin are inhibited by N-acetlyneuraminic acid. Homologous proteins in other oral streptococci not only bind to salivary agglutinins (16, 36, 40, 95, 203) but also to other components of the salivary pellicle (101, 132), and they can mediate coaggregation with A. naeslundii (91) and P. gingivalis (125).

RECEPTORS

Many viridans streptococci synthesize cell wall polysaccharides that consist of a strain-specific hexa- or heptasaccharide repeating unit linked end to end by phosphodiester bonds (Table 3). Cisar, Bush, and their collaborators have purified and structurally identified several polysaccharides (1-4, 170) that are the likely receptors for the

Gal/GalNAc-reactive lectins of streptococci (32, 113,[cont. on p.532] 215), veillonellae (82), prevotellae (140), hemophilli (121), and those associated with the type 2 fimbriae on oral actinomyces (29, 30, 99). Cassels and coworkers purified from S. oralis ATCC 55229 a rhamnose-containing polysaccharide that inhibited coaggregation between the streptococcus and a capnocytophaga (20, 22), an L-rhamnose-inhibitable coaggregation. This repeating hexasaccharide does not contain GalNAc residues and is not linked by phosphodiester bonds, but rather the phosphate is found as a glycerol phosphate substitution on an internal galactose moiety (69). The polysaccharide from S. mitis K103 (171) does not have the GalNAc b-(1 --> 3) Gal or Gal b-(1 --> 3)GalNAc found in the top five polysaccharide sequences (Table 3, underlined disaccharide), and strain K103 exhibits a different coaggregation pattern (81).

The GalNAc b-(1 --> 3)Gal or Gal b-(1 --> 3) GalNAc disaccharides are part of glycoconjugate receptors found on the surface of human cells. Oral bacteria may evade the host secretory immune response by expressing these glycoconjugates. Other bacteria like actinomyces and veillonellae may mask their presence by surrounding their cell bodies through coaggregations with these streptococci. Rosettes and corncob morphological arrangements composed of one cell type surrounded by a different cell type are common in dental plaque (103, 104). The fine specificity of oral actinomyces lectins observed in their adherence to immobilized glycolipids with exposed GalNAc b-containing glycoconjugates has been proposed to reflect the actinomyces tissue-specific colonization (195, 196). In a study of actinomyces coaggregation with S. oralis 34 [GalNAc b-(1 --> 3) Gal-containing polysaccharide, Table 3] and S. mitis J22 [Gal b-(1 --> 3) GalNAc-containing polysaccharide, Table 3], only a minor difference in specificity was noted (30). However, a major difference was seen in the ability of streptococci and actinomyces to recognize these two disaccharides (30). Whereas the actinomyces coaggregate with both S. oralis 34 and S. mitis J22 and exhibit only a twofold difference in inhibition by either disaccharide, intrageneric coaggregations between S. oralis 34 [GalNAc b-(1 --> 3) Gal-containing polysaccharide, Table 3] and S. gordonii DL1 (expresses 100-kDa putative adhesin; 32, 215) are 30-fold more inhibited by GalNAc b-(1 --> 3) Gal (30). Cisar et al (30) proposed that the actinomyces recognize the sides of the disaccharides that are structurally similar, while the streptococci recognize the side of the GalNAc b-(1 --> 3) GAl that contains the acetamido group. Such clear specificity is likely to have a role in the complex microbial interactions that construct oral biofilms.

Molecular modeling of the streptococcal polysaccharides of S. oralis 34, S. oralis J22 (also called S. mitis J22) and S. oralis ATCC 55229 indicated that the polysaccharide can fold to form a loop initiated at the Galf and consisting of three or four sugar residues. The loop is stabilized by hydrogen bonding interactions between the nearest phosphate to the reducing end and the Galf residue, which provides the needed flexibility (21). Within the loop is the lectin-binding site (Table 3, underlined disaccharides). Since GalNAc b-(1 --> 3)Gal-containing capsular polysaccharide seems to be the streptococcal adhesin-binding site, it is not surprising that streptococci, e.g. S. oralis J22 and S. oralis ATCC 55229, lacking this sequence do not participate in intrageneric coaggregation (113). The cell wall polysaccharide of S. mitis K103 (Table 3) contains no galactofuranose and no lectin-binding site, and its coaggregation properties are distinct from nearly all the other 71 strains of viridans streptococci examined in a recent survey (81).

ACTINOMYCES ADHERENCE

The most-studied actinomyces are members of A. naeslundii (many were formerly called Actinomyces viscosus) (92), Actinomyces israelii, Actinomyces odontolyticus, and Actinomyces serovar WVA963. As with the streptococci, actinomyces express multiple kinds of adherence mechanisms, which include fimbriae-attached proteins and protease-resistant receptors recognized by adhesins on cells of other genera (103).

Two kinds of fimbriae have been reported. Type 1 fimbriae mediate binding to SHA and specifically to salivary proline-rich proteins. Type 2 fimbriae mediate lactose-inhibitable binding to other oral bacteria and to mammalian cells. Some actinomyces strains bear both types of fimbriae, and

others bear only type 2. No natural isolates bearing only type 1 fimbriae have been reported, although mutants have been constructed that express only type 1 fimbriae.

The genes encoding the fimbrial subunits of type 1 and type 2 fimbriae of A. naeslundii strains T14V (type 1+, type 2+) and WVU45 (type 1-, type 2+) have been cloned and sequenced (43, 221-223). Both subunits have a molecular size of about 60 kDa, and both contain amino terminal signal sequences and the cell surface binding motif LPxTG in the carboxy-terminal region that is characteristic of many gram-positive surface proteins (184, 185). The deduced amino acid sequences of A. naeslundii T14V FimA (fimA gene product; structural subunit of type 2 fimbriae) and that of A. naeslundii WVU45 FimP (fimP gene product; structural subunit of type 1 fimbriae) are 34[percent] identical (223). The amino acid sequence identity between the FimA proteins of A. naeslundii WVU45 and A. naeslundii T14V is 70[percent] (222; JA Donkersloot, personal communication). However, it remains unclear if the FimA subunit possesses the lactose-sensitive adhesive function or if an accessory fimbrial protein is responsible (28).

A significant advance in studies of actinomyces adherence has been made by Yeung and colleagues, who reported a genetic transfer system (224) and used integration plasmids to generate site-specific mutations in A. naeslundii (220). In a careful study of three such mutants, it was determined that fimP was essential for type 1 fimbrial synthesis, but that fimP was not sufficient for conferring bacterial adherence (220). The 3' region near fimP on the A. naeslundii T14V chromosome may be the site of the putative adhesin gene. A minor protein at the tip of the type 1 fimbriae possibly mediates the binding of A. naeslundii T14V to salivary proline-rich proteins (162).

Actinomyces mutants missing type 2 fimbriae fail to coaggregate with their streptococcal partners. Our laboratory and collaborators initiated a study of Actinomyces serovar WVA963 strain PK1259, the reference strain of the actinomyces coaggregation group F (117). This strain was chosen because it exhibits only coaggregation that is lactose-inhibitable with reference strains of three oral streptococcal coaggregation groups (102). Considering that these coaggregations may all be mediated by type 2 fimbriae, coaggregation defective (COG-) mutants should be lacking these fimbriae. Although Actinomyces serovar WVA 963 is among the top 20 most frequently isolated bacteria from subgingival sites, nothing has been reported about its fimbrial structures.

Two coaggregation phenotypes were observed among the mutants, and these are depicted diagrammatically in Figure 2. On the left side is shown the wildtype strain PK1259 and its coaggregations with V. atypica PK1910 (complementary set of triangles and triangles with stem), F. nucleatum PK1594 (complementary set of semicircles and semicircles with stem), and the members of streptococcal coaggregation groups 3, 4, and 5 (complementary sets of rectangles and rectangles with stems). No inhibitors are known for the interactions with veillonella and fusobacteria. Both the wild type and COG- mutants exhibit fimbriae (100). Type 1 and type 2 fimbriae are composed of subunits of 59 kDa and 57 kDa, respectively (100). The type 2 fimbriae are depicted in Figure 2 as protruding downward and bearing at the tip the proposed lactosesensitive adhesins that mediate coaggregations with the three groups of streptococci. Type 1 fimbriae are represented as protruding from the top of the cell, and they are also present on COG- mutant strains PK2407 and PK3092, the two strains representing the two mutant phenotypes. All three strains bind to salivary proline-rich protein-coated latex beads, produce the 59-kDa subunit of type 1 fimbriae, and are agglutinated by antiserum against A. naeslundii type 1 fimbriae (100). COG- mutant PK2407 retains coaggregation with F. nucleatum PK1594 but does not coaggregate with V. atypica PK1910, whereas COG- mutant PK3092 retains coaggregation with both the fusobacterium and the veillonella.

Antisera against the wild type PK1259 that have been absorbed with the COG- mutant PK3092 blocked coaggregations between the wild type and the[cont. on p.536] three streptococcal partners (100). Surprisingly, antisera absorbed with COG- mutant PK2407 removed the blocking **antibodies**. Antiserum against A. naeslundii T14V type 2 fimbriae agglutinated the wild type and COG- mutant PK3092, but it did not agglutinate COG- mutant PK2407. Anti-type 2 serum reacted with a 57-kDa protein on immunoblots of surface proteins of the wild type and COG- mutant PK3092, but this protein

was absent from COG- mutant PK2407 (100). These results suggested that a separate protein from the type 2 fimbrial subunit was responsible for the lactose-inhibitable coaggregation with streptococci. Wild-type cells subjected to mild sonication released a 95-kDa protein that specifically bound to lactose-agarose beads and was eluted by lactose (100). COG- mutant PK3092 did not have this protein. In contrast, larger amounts of the 95-kDa protein were released from COG- mutant PK2407 than from the wild type (100). These observations are depicted in Figure 2 as improper orientation of putative lactose-sensitive adhesin (solid rectangles) on the PK2407 cell surface; the orientation is opposite to that required for proper functioning as a coaggregation mediator. However, since COG- mutant PK3092 bears the 57-kDa type 2 fimbrial subunit but not the 95-kDa putative adhesin, it is proposed that Actinomyces serovar WVA963 PK1259 mediates lactose-inhibitable coaggregation with streptococci by a 95-kDa lactose-sensitive adhesin that is a minor protein attached to the type 2 fimbriae.

The lactose-inhibitable coaggregations between actinomyces adhesins and streptococcal cognate receptors were reported to have a much greater impact than nonspecific interactions in a study of actionmyces binding to streptococcal-coated hexadecane droplets (50). The results also agree with the independent nature of the multiple kinds of possible interactions by a community of oral bacteria (107). For example, lactose-inhibitable coaggregations among cells in dental plaque are **prevented** in the presence of lactose, but lactose-noninhibitable interactions bind the community as a multigeneric coaggregate. An analysis of the kinetics of accretion of streptococci onto actinomyces bound to the substratum in a parallel flow chamber indicated a 19-fold higher accretion with coaggregating bacteria than with a COG- mutant of the streptococcus (17). The results are best explained by collisions occurring between the parallelflowing, accreting streptococci and the already adherent actinomyces.

Actinomyces naeslundii WVU45 cells have been shown to bind to human buccal epithelial cells by recognizing a 180-kDa salivary glycoprotein that was bound to the epithelial cells (12). In a different study, both a 130-kDa glycoprotein and certain gangliosides were implicated in the attachment of A. naeslundii WVU45 to sialidase- **treated** polymorphonuclear leukocytes (178). This actinomyces, as well as numerous others, possesses potent sialidases, and the sialidase genes from the various actinomyces are highly conserved (219).

PREVOTELLA ADHERENCE

Prevotella loescheii PK1295 bears a galactoside-specific adhesin that mediates hemagglutination of neuraminidase- **treated** human erythrocytes (212) and lactose-inhibitable coaggregations with oral streptococci (118). The adhesin is a 75-kDa protein (140) that is expressed on the cell surface in a maximum of 400 molecules per cell (214). The adhesin is associated with the distal portion of the bacterium's fimbriae. The gene plaA (prevotella loescheii adhesin A) (146) contains a programmed frameshifting hop (147), an infrequently found event in prokaryotes. In such events the ribosome reads through the interrupted sequence and realigns itself in the correct reading frame for translating the remainder of the message. The role of a frameshifting hop in plaA is unclear. Perhaps the hop allows a coupling of two functions, of which the second is yet to be discovered. P. loescheii PK1295 expresses a second adhesin of 45 kDa, which mediates lactose-noninhibitable coaggregation with A. israelii PK14 (213). By using **monoclonal antibodies** to the 75- and 45-kDa adhesins, it was shown by immunoelectronmicroscopy that essentially all cells in a population express both adhesins (141).

Earlier studies had shown that the coaggregation partners of prevotellae appear to be primarily actinomyces (103), and two recent surveys confirmed these data and showed that the prevotella bear the protein coaggregation adhesin and the actinomyces bear a carbohydrate-containing receptor on their respective cell surface (34, 163). The actinomyces partners are mostly members of actinomyces coaggregation groups C, D, and E (102). The prevotellae also coaggregate with F. nucleatum and in most of the coaggregations, the fusobacterium is the heat-inactivated cell type (112). The prevotellae synthesize proteases and one of these synthesized by P. loescheii degraded a 75-kDa PlaA adhesin (23). The authors proposed that the protease may actually aid in

detachment of the prevotellae from the coaggregation partner cell and, thus, allow the prevotellae to search out a different econiche in plaque. This attractive idea follows the notion that bacterial community members possess sensory surface molecules that may couple sensing to protease activity, which activates detachment strategies and initiates the search for a more suitable econiche (104).

PORPHYROMONAS ADHERENCE

P. gingivalis cells exhibit a broad capability of binding to various oral surfaces such as other oral bacteria (70, 97, 108, 112, 126, 135, 157, 173, 193, 194, 218), salivary components (63, 145), fibronectin-collagen complexes (160), erythrocytes and monocytes (166), and epithelial cells (25, 85). A 150-kDa surface component of P. gingivalis binds to fibrinogen (131). Fibrinogen (158) and histatins (155) inhibit coaggregation between porphyromonads and streptococci.

Coaggregation between P. gingivalis and S. gordonii G9B is inhibited by a 43-kDa salivary protein (193), but proteases rapidly degraded this protein. Both porphyromonads (5, 131) and streptococci (128, 138) produce proteases which may function in eliminating inhibitors. In contrast, surface proteases of P. gingivalis may mediate coaggregations with A. viscosus (135). Considering the proposed role of proteases in inactivating adhesins in prevotellae (23), the role of proteases in porphyromonads adherence may be different. In addition, the fimbrial gene fimA of P. gingivalis could be insertionally inactivated without affecting the ability to coaggregate with S. gordonii G9B, but adherence to SHA was reduced in the mutant (145). In a different study, purified fimbriae of P. gingivalis bound to S. gordonii G9B and blocked the porphyromonas-streptococcus coaggregation (122). The authors concluded that the fimbriae were at least partly responsible for coaggregation with S. gordonii G9B (122). Purified porphyromonas fimbriae were also shown to bind to A. naeslundii (70). So, the porphyromonad fimbriae may play a role in coaggregation, but no minor accessory protein with an adhesin function has been reported.

P. gingivalis also produces extracellular vesicles that bind to serum-, saliva-, and crevicular fluid-coated hydroxyapatite (27), mediate binding of streptococci to serum-coated hydroxyapatite (189), and act as coaggregation bridges between the noncoaggregating Eubacterium saburreum and Capnocytophaga ochracea (73). The observation that vesicles could bind to fibronectin, fibrinogen, collagen, and laminin (45) suggests that P. gingivalis may employ these vesicles to enhance their advantage in colonizing oral surfaces.

FUSOBACTERIUM ADHERENCE

As a group, strains of Fusobacterium nucleatum coaggregate with members of all oral bacterial genera so far examined. These include 12 genera surveyed in 1989 (112) and Treponema, as well as intrageneric coaggregations among F. periodonticum and F. nucleatum strains (120). The first report of the extensive nature of lactose-inhibitable coaggregations among gram-negative partnerships indicated that, of 43 gram-negative pairs, 22 were completely inhibited and another 14 were partially inhibited by 60 mM lactose (112). Further studies with F. nucleatum and Porphyromonas gingivalis showed equal inhibition with lactose, N-acetyl-D-galactosamine, and D-galactose (108). The fusobacterium was inactivated by heat- or protease- **treatment** , but the porphyromonas cells were unaffected by these **treatments** (108). These results were confirmed and extended by demonstrating that the porphyromonad was inactivated by sodium metaperiodate **treatment** (97). Also, antiserum against a 42-kDa outer membrane protein of F. nucleatum blocks coaggregation with P. gingivalis (98). Coaggregation-defective mutants of F. nucleatum selected for inability to coaggregate with P. gingivalis also could no longer exhibit lactose-inhibitable coaggregation with other gram-negative partners (N Ganeshkumar & PE Kolenbrander, unpublished data). Collectively, these data suggest that coaggregation between F. nucleatum and several gram-negative partners may be mediated by the same galactoside-sensitive fusobacterial surface component.

F. nucleatum binds to an 89-kDa salivary proline-rich glycoprotein, and deglycosylation of the glycoprotein by b-galactosidase abolishes binding (67). Adherence of F. nucleatum to human peripheral blood

lymphocytes is blocked by galactosides (204). Fibronectin is a component of saliva, and fusobacteria bind to epithelial cells coated with fibronectin (13). This binding can be inhibited by coating the fusobacteria with fibronectin, laminin, and type IV collagen. A fusobacterial hemagglutinin aggregated 14 of 17 strains of oral streptococci, and the hemagglutination was inhibited by L-arginine (199). A polypeptide of 39.5 kDa was obtained from the surface of F. nucleatum, and it inhibited coaggregation of the fusobacterium with streptococcus (94). Antiserum against the polypeptide blocked coaggregation. In a survey of 33 oral bacterial strains, only F. nucleatum was a coaggregation partner of five Eubacterium species (62). Arginine, histidine, lysine, and glycine inhibited these coaggregations. L-Arginine also inhibits the interaction between a hemagglutinin from F. nucleatum and oral streptococci (199). Elucidation of the wide variety of options for adherence by oral fusobacteria helps to explain why these bacteria are among the most numerous bacteria in dental plaque.

TREPONEMA ADHERENCE

Some coaggregations between T. denticola and fusobacteria are inhibited by D-galactosamine (120). A survey of 22 strains of Treponema spp., including all four named human oral species, showed that all coaggregated with selected F. nucleatum and F. periodonticum strains but not with members of 9 other genera (120). While the fusobacteria were inactivated by heat **treatment** , the treponemes were resistant to heating. In a separate study both treponemes and P. gingivalis partner cells required heating to eliminate coaggregation (71). Interactions of T. denticola with P. gingivalis were partially inhibited by saliva and serum (218). It is not surprising that T. denticola and P. gingivalis are found together in deep periodontal pockets (96). Moreover, T. denticola was never found in periodontically affected sites unless P. gingivalis was also present (188). Bacteroides forsythus and S. crista coaggregate with T. denticola, and only the interaction with S. crista is inhibited by saliva (218). An interesting model of polar adhesion of T. denticola to surfaces has been proposed (49).

VEILLONELLA ADHERENCE

Veillonellae are metabolically poised to flourish in communities dominated by streptococci and actinomyces because they utilize the lactic acid end products produced during the growth of the latter two groups on sugars. The veillonellae are detectable shortly after streptococci colonize the sterile neonate oral cavity. Some streptococci coaggregate with Veillonella atypica PK1910 by lactose-inhibitable coaggregations, but a class of V. atypica mutants has completely lost the ability to participate in these coaggregations (83). Both parent and mutant cells exhibit fimbriae, but only the parent produces a 45-kDa surface protein that binds to streptococcal cells and to lactose-agarose beads (82). Antiserum prepared against the 45-kDa protein-bound lactose-agarose beads blocked veillonella-streptococcus coaggregations. It appears that the 45-kDa putative adhesin mediates lactose-inhibitable coaggregation and that it is not likely to be a structural subunit of the fimbriae.

CAPNOCYTOPHAGA ADHERENCE

Capnocytophaga are gram-negative bacteria associated with moderate periodontal lesions (153, 169). They do not possess fimbriae; their adhesin molecules are presumably intercalated into the outer membrane. Tempro et al found in their experiments that a 140-kDa polypeptide from the outer membrane of Capnocytophaga gingivalis DR2001 was recognized by **monoclonal antibodies** that blocked coaggregation between the capnocytophaga and Actinomyces israelii PK16 (201). Radio-iodinated **monoclonal antibodies** were used to calculate values of between 220 and 280 adhesin sites per cell. The sites were arranged nonuniformly on the cell surface.

In experiments by Weiss et al, a rhamnose-sensitive adhesin from C. ochracea ATCC 33596 was determined to be 155 kDa by using **monoclonal antibodies** that blocked coaggregation with S. oralis ATCC 55229 (see Table 3 for structure of receptor) (211). They proposed that the adhesin is an

outer membrane protein.

INVASION

Several bacteria isolated from the oral cavity have been seen inside epithelial cells. These include T. denticola (49), P. gingivalis (127), and Actinobacillus actinomycetemcomitans (152). Invasion is considered an important virulence factor, offering protection from the host immune system and contributing to tissue damage in a nutritionally rich environment that is free of competing organisms. The bacteria first attach to the epithelial cell membrane and then induce a series of structural and biochemical changes that facilitate bacterial penetration.

TREPONEMA DENTICOLA

Confocal microscopy and transmission electron microscopy revealed T. denticola fully within human gingival fibroblasts that are close to and in the same plane as the nucleus and wound around actin filaments (49). Other studies indicate that a T. denticola cell has been detected only rarely intracellularly (38). By using migrating epithelial cells, it has been shown that when T. denticola is added, it leads to extensive membrane blebbing of some solitary cells (205). Following longer exposure to the bacteria, some cells show large blebs, membrane damage, and invasion of T. denticola.

T. denticola produces a chymotrypsin-like outer membrane-associated enzyme. The purified proteinase causes membrane blebbing, perhaps by degrading the connection of actin to the cell membrane (205). Diffusion of the proteinase or other toxic substances inside the epithelial cells may interfere with the intracellular signalling networks and lead to blebbing. Most of the surface cells of epithelial multilayers bind few, if any, T. denticola cells and have shown little blebbing as evidenced by scanning electron microscopy. Neighboring cells have been shown to retract on exposure to bacteria. Transmission electron microscopy has showed no evidence of intact T. denticola inside the epithelial multilayers; however, immunogold electron microscopy has indicated that the bacterial antigens, including the chymotrypsin-like proteinase, penetrate into the cell cytoplasm and accumulate inside the intracellular vacuoles. The chymotrypsin-like proteinase is associated with a 53-kDa abundant outer membrane treponema protein. This protein complex seems to play a role in the attachment and virulence of the organism. The protein mediates attachment to fibronectin, laminin, and fibrinogen (74) and is able to create unusually large pores in artificial lipid membranes (48).

PORPHYROMONAS GINGIVALIS

Bacterial colonization of gingival tissue and its penetration and destruction are critical processes in the pathogenesis of periodontal disease. P. gingivalis adheres in vitro to human buccal epithelial cells and gingival sulcular epithelial cells. Electron microscopy demonstrated that P. gingivalis adheres to and internalizes within primary cultures of gingival epithelial cells (127). The epithelial cells showed no visible damage. P. gingivalis binds to and invades multilayered gingival pocket epithelial cells (180) and a human oral epithelial cell line (KB) (46, 179). Internalized P. gingivalis is found associated directly with the epithelial cell cytoplasm and not encapsulated by endocytic vacuoles (123, 180). Lamont et al (123) found that invasion was significantly lower during the lag phase of growth. Protein synthesis is necessary, and about 90 min is required to complete the invasion process for most of the bacteria. Increasing the incubation of bacteria with epithelial cells to 2 h did not result in increased invasion. However, killing of the external bacteria after 90 min of incubation, followed by an additional 4 h of incubation, resulted in a higher recovery of internal P. gingivalis. These results indicate that the bacteria are dividing within the epithelial cells.

ACTINOBACILLUS ACTINOMYCETEMCOMITANS

Several mechanisms have been identified for adherence of A. actinomycetemcomitans to epithelial cells (56). Extracellular amorphous material, fimbriae, and extracellular vesicles may all be involved (150,

151). Certain A. actinomycetemcomitans strains undergo a variant shift that is associated with changes in adherence and invasion properties. Smooth variants invade more proficiently than rough variants. Fimbriae most probably function in adherence of rough variants, whereas non-fimbrial components (eg. vesicles) are probably involved in adherence of smooth, highly invasive strains (152). Transmission electron microscopy of the cell monolayers has revealed A. actinomycetemcomitans within the KB cells. The bacteria were surrounded by an endosomal vacuole of the KB cell that disintegrates, releasing the organisms into the cytoplasm (192). Invasion is an active process; both the KB cells and the bacteria must be metabolically active. Accordingly, the likelihood of invasion is much higher with midexponential phase cells.

BACTERIAL COMMUNICATION

Physical adherence of cells is communication. The recent evidence for involvement of lipoproteins in adherence as well as in substrate binding suggests that surface localized binding proteins may act as communicators to assist or perhaps to mediate permanent cell-to-cell contact. The enterococcal pheromone binding lipoproteins PrgZ (174) and TraC (200) also illustrate adherence-relevant functions in bacterial communication. Furthermore, a signal peptide fragment from a staphylococcal protein, Trah, exhibits enterococcal pheromone activity (55). Such examples of breadth in communication among microbes suggest that numerous mechanisms of communication exist in a biofilm like dental plaque. We have examined (PE Kolenbrander & EP Greenberg, unpublished data) the possibility that autoinducer homologs of the homoserine lactones (58) are produced by gram-negative oral bacteria and by cultures of unaltered dental plaque scrapings. No homologs were detected by three assay systems, suggesting that different communication signals are exchanged than those from homoserine lactones.

Biofilms representing several econiches have been studied, but the development and use of the scanning confocal laser microscope has been a critical advance in our ability to monitor these biofilm communities. SCLM allows nondestructive experimentation and has just recently been applied to oral communities. The same technology has been exploited in a study of a biofilm that degraded a commercial herbicide (217). The investigators observed highly specific patterns of intra- and intergeneric coaggregations. They also suggest that adaptation of a biofilm to degrade a recalcitrant substrate involves restructuring the community and development of different cellular arrangements to promote efficient metabolic communication (217).

Added material

Catherine J. Whittaker, Christiane M. Klier, and Paul E. Kolenbrander
Laboratory of Microbial Ecology, National Institute of Dental Research, National Institutes of Health, Bethesda, Maryland 20892

Table 1 Frequently isolated human oral bacteria and some adherence properties

Genus	Coaggregation	Invasion
Actinobacillus	(112)	(152)
Actinomyces	(66)	--
Bacteroides(FNa)	(218)	--
Campylobacter(FNb)	(118)	--
Capnocytophaga	(114)	--
Corynebacterium (formely called Bacterionema)	(154)	--
Eikenella	(47)	--
Eubacterium	(62)	--
Fusobacterium	(112)	--
Gemella	(112)	--
Haemophilus	(137)	--
Lactobacillus	(216)	--
Neisseria	(66)	--
Peptostreptococcus	(112)	--
Porphyromonas (formerly called Bacteroides in part)	(112)	(127)
Prevotella	(111)	--

(formerly called Bacteroides in part)		
Propionibacterium	(26)	--
Rothia	(106)	--
Selenomonas	(112)	--
Streptococcus	(66)	--
Treponema	(120)	(205) (FNc)
Veillonella	(66)	--

FOOTNOTES

a Only a few species including Bacteroides forsythus remain. Most are reclassified as either Prevotella or Porphyromonas species.

b Genus contains strains previously classified as Wolinella.

c Transport of T. denticola chymotypsin-like proteinase into newly formed large intracellular vacuoles within the epithelial cells was reported.

Table 2 Adhesins and binding proteins of human oral bacteria(FNa)
{Table omitted}

FOOTNOTE

a Former names: Actinomyces naeslundii (A. viscosus) T14V; Actinomyces naeslundii (A. viscosus) WVU627; Porphyromonas gingivalis (Bacteroides gingivalis) ATCC 33277; Prevotella loescheii (Bacteroides loescheii) PK1295; Streptococcus crista (S. sanguis) CC5A; Streptococcus gordonii (S. sanguis); Streptococcus oralis ATCC 55229 (S. sanguis H1); Streptococcus parasanguis (S. sanguis) FW213; Veillonella parvula (V. alcalescens) V1

Table 3 Polysaccharide receptors of viridans streptococci(FNa)
{Table omitted}

FOOTNOTE

a Abbreviations: Rha, L-Rhamnose; Gal, D-Galactose; Glc, D-Glucose; GalNAc, N-Acetyl-D-Galactosamine; Glyc, Glycerol; p, pyranose; f, furanose. Square brackets indicate repeating hexa- to octa-saccharide units. Underlined disaccharides indicate lectin-binding sites.

Figure 1 Diagrammatic representation of streptococcal adherence to other streptococci, to other genera of oral bacteria, and to the acquired pellicle coating on teeth, all depicted as complementary pairs of symbols. Putative adhesin, obelisk with stem; complementary receptor, obelisk. The obelisk with stem represents a cell surface molecule known to be inactivated by heat (85[degree]C/30 min), whereas its complementary symbol is insensitive to the same amount of heating. Inset box at center right: the obelisk and ellipse represent lactose-noninhibitable interactions; the bottom two rectangles represent lactose-inhibitable coaggregations. The light gray rectangle represents coaggregation between Streptococcus SM PK509 and Streptococcus oralis 34 (113), shown in the upper left corner. Solid black rectangles indicate coaggregation with S. gordonii PK488 (118) and are shown as mediators of several other coaggregations. See text for more information.

Figure 2 Diagrammatic representation of the role of actinomyces type 2 fimbriae and the putative adhesin in mediating coaggregation of Actinomyces serovar WVA963 strain PK1259 with oral streptococcus coaggregation groups 3, 4, and 5. Coaggregation-defective mutant strains PK2407 and PK3092 are shown, lacking type 2 fimbriae and putative adhesin, respectively. Type 1 fimbriae are represented as zigzag lines. The type 2 fimbriae are represented by solid lines. Lactose-inhibitable coaggregations are depicted by rectangles, as in Figure 1. See text for more description. Also represented is the independent nature of interactions between Actinomyces serovar WVA963 and its coaggregation partner streptococci, fusobacteria, and veillonellae.

ACKNOWLEDGMENTS

We thank J Cisar, M Curtis, P Fives-Taylor, C Hughes, H Jenkinson, and R Lamont for providing prepublication manuscripts, and we thank R Andersen, J London, and S Robrish for helpful comments during the writing of this review.

LITERATURE CITED

1. Abeygunawardana C, Bush CA, Cisar JO. 1990. Complete structure of the polysaccharide from Streptococcus sanguis J22. Biochemistry 29:234-48

2. Abeygunawardana C, Bush CA, Cisar JO. 1991. Complete structure of the cell surface polysaccharide of Streptococcus oralis ATCC 10557: a receptor for lectinmediated interbacterial adherence. Biochemistry 30:6528-40

3. Abeygunawardana C, Bush CA, Cisar JO. 1991. Complete structure of the cell surface polysaccharide of Streptococcus oralis C104: a 600-MHz NMR study. Biochemistry 30:8568-77

4. Abeygunawardana C, Bush CA, Tjoa SS, Fennessey PV, McNeil MR. 1989. The complete structure of the capsular polysaccharide from Streptococcus sanguis 34. Carbohydr. Res. 191:279-93

5. Aduse-Opoku J, Muir J, Slaney JM, Rangarajan M, Curtis MA. 1995. Characterization, genetic analysis, and expression of a protease antigen (PrpRI) of Porphyromonas gingivalis W50. Infect. Immun. 63:4744-54

6. Ali RW, Skaug N, Nilsen R, Bakken V. 1994. Microbial associations of 4 putative periodontal pathogens in sudanese adult periodontitis patients determined by DNA probe analysis. J. Periodontol. 65:1053-57

7. Alloing G, dePhilip P, Claverys JP. 1994. Three highly homologous membranebound lipoproteins participate in oligopeptide transport by the Ami system of the gram-positive Streptococcus pneumoniae. J. Mol. Biol. 241:44-58

8. Alloing G, Trombe M-C, Claverys JP. 1990. The ami locus of the grampositive bacterium Streptococcus pneumoniae is similar to binding proteindependent transport operons of gramnegative bacteria. Mol. Microbiol. 4:633-644

9. Ames GF-L. 1992. Bacterial periplasmic permeases as model systems for the superfamily of traffic ATPases, including the multidrug resistance protein and the cystic fibrosis transmembrane conductance regulator. Int. Rev. Cytol. 137A:1-35

10. Andersen RN, Ganeshkumar N, Kolenbrander PE. 1993. Cloning of the Streptococcus gordonii PK488 gene, encoding an adhesin which mediates coaggregation with Actinomyces naeslundii PK606. Infect. Immun. 61:981-87

11. Andersen RN, Lunsford RD, Kolenbrander PE. 1995. Transcript size of putative sca operon of Streptococcus gordonii PK488. In Abstracts of the 95th General Meeting of the American Society for Microbiology 1995, Abstr. D-66, pp. 260. Washington, DC: Am. Soc. Microbiol.

12. Babu JP, Dabbous MK, Abraham SN. 1991. Isolation and characterization of a 180-kiloDalton salivary glycoprotein which mediates the attachment of Actinomyces naeslundii to human buccal epithelial cells. J. Periodont. Res. 26:97-106

13. Babu JP, Dean JW, Pabst MJ. 1995. Attachment of Fusobacterium nucleatum to fibronectin immobilized on gingival epithelial cells or glass coverslips. J. Periodontol. 66:285-290

14. Bartsevich VV, Pakrasi HB. 1995. Molecular identification of an **ABC transporter** complex for manganese: analysis of a cyanobacterial mutant strain impaired in the photosynthetic oxygen evolution process. EMBO J. 14:1845-1853

15. Blanco DR, Champion CI, Exner MM, Erdjument-Bromage H, Hancock REW, et al. 1995. Porin activity and sequence analysis of a 31-kilodalton Treponema pallidum subsp. pallidum rare outer membrane protein (Tromp 1). J. Bacteriol. 177:3556-62

16. Bleiweis AS, Oyston PCF, Brady LJ. 1992. Molecular, immunological and functional characterization of the major surface adhesin of Streptococcus mutans. In Genetically Engineered Vaccines, ed. JE Ciardi, JR McGhee, JM Keith, pp. 229-41. New York: Plenum

16a. Bloomquist CG, Reilly BE, Liljemark WF. 1996. Adherence, accumulation, and cell division of a natural adherent baterial population. J. Bacteriol. 178:1172-77

17. Bos R, van der Mei HC, Busscher HJ. 1995. A quantitative method to study coadhesion of microorganisms in a parallel plate flow chamber. II. Analysis of the kinetics of co-adhesion. J. Microbiol. Meth. 23:169-82

18. Brady LJ, Piacentini DA, Crowley PJ, Oyston PC, Bleiweis AS. 1992. Differentiation of salivary agglutinin-mediated adherence and aggregation of mutans streptococci by use of **monoclonal antibodies** against the major surface adhesin P1. Infect. Immun. 60:1008-17

19. Burnette-Curley D, Wells V, Viscount H, Munro CL, Fenno JC, et al.

1995. FimA, a major virulence factor associated with Streptococcus parasanguis endocarditis. Infect. Immun. 63:4669-74

20. Cassels FJ, Fales HM, London J, Carlson RW, van Halbeek H. 1990. Structure of a streptococcal adhesin carbohydrate receptor. J. Biol. Chem. 265:14127-35

21. Cassels FJ, Hughes CV, Nauss JL. 1995. Adhesin receptors of human oral bacteria and modeling of putative adhesin-binding domains. J. Ind. Microbiol. 15:176-85

22. Cassels FJ, London J. 1989. Isolation of a coaggregation-inhibiting cell wall polysaccharide from Streptococcus sanguis H1. J. Bacteriol. 171:4019-25

23. Cavedon K, London J. 1993. Adhesin degradation: a possible function for a Prevotella loescheii protease? Oral Microbiol. Immunol. 8:283-87

24. Chandad F, Mouton C. 1995. Antigenic, structural, and functional relationships between fimbriae and the hemagglutinating adhesin HA-Ag2 of Porphyromonas gingivalis. Infect. Immun. 63:4755-63

25. Childs WC III, Gibbons RJ. 1990. Selective modulation of bacterial attachment to oral epithelial cells by enzyme activities associated with poor oral hygiene. J. Periodont. Res. 25:172-78

26. Ciardi JE, McCray GFA, Kolenbrander PE, Lau A. 1987. Cell-to-cell interaction of Streptococcus sanguis and Propionibacterium acnes on saliva-coated hydroxyapatite. Infect. Immun. 55:1441-46

27. Cimasoni G, Song M, McBride BC. 1987. Effect of crevicular fluid and lysosomal enzymes on the adherence of streptococci and bacteroides to hydroxyapatite. Infect. Immun. 55:1484-89

28. Cisar JO, Barsumian EL, Siraganian RP, Clark WB, Yeung MK, et al. 1991. Immunochemical and functional studies of Actinomyces viscosus T14V type 1 fimbriae with **monoclonal** and polyclonal **antibodies** directed against the fimbrial subunit. J. Gen. Microbiol. 137:1971-79

29. Cisar JO, Curl SH, Kolenbrander PE, Vatter AE. 1983. Specific absence of type 2 fimbriae on a coaggregation-defective mutant of Actinomyces viscosus T14V. Infect. Immun. 40:759-65

30. Cisar JO, Sandberg AL, Abeygunawardana C, Reddy GP, Bush CA. 1995. Lectin recognition of host-like saccharide motifs in streptococcal cell wall polysaccharides. Glycobiology 5:655-62

31. Clark WB, Beem JE, Nesbitt WE, Cisar JO, Tseng CC, Levine MJ. 1989. Pellicle receptors for Actinomyces viscosus type 1 fimbriae in vitro. Infect. Immun. 57:3003-8

32. Clemans DL, Kolenbrander PE. 1995. Identification of a 100-kDa putative coaggregation-mediating adhesin of Streptococcus gordonii DL1 (Challis). Infect. Immun. 63:4890-93

33. Clemans DL, Kolenbrander PE. 1995. Isolation and characterization of coaggregation-defective (Cog-) mutants of Streptococcus gordonii DL1 (Challis). J. Ind. Microbiol. 15:193-97

34. Cookson AL, Handley PS, Jacob AE, Watson GK, Allison C. 1995. Coaggregation between Prevotella nigrescens and Prevotella intermedia with Actinomyces naeslundii strains. FEMS Microbiol. Lett. 132:291-96

35. Correia F, DiRienzo JM, McKay T, Rosan B. 1995. Cloning and nucleotide sequence analysis of a new member of the streptococcal 37-kDa adhesin family. J. Dent. Res. (special issue) 74:200

36. Crowley PJ, Brady LJ, Piacentini DA, Bleiweis AS. 1993. Identification of a salivary agglutinin-binding domain within cell surface adhesin P1 of Streptococcus mutans. Infect. Immun. 61:1547-52

37. Cundell DR, Pearce BJ, Sandros J, Naughton AM, Masure HR. 1995. Peptide permeases from Streptococcus pneumoniae affect adherence to eucaryotic cells. Infect. Immun. 63:2493-98

38. De Filippo **AB** , Ellen RP, McCulloch CAG. 1995. Induction of cytoskeletal rearrangements and loss of volume regulation in epithelial cell by Trepenoma denticola. Arch. Oral Biol. 40:199-207

39. de Jong MH, van der Hoeven JS. 1987. The growth of oral bacteria on saliva. J. Dent. Res. 66:498-505

40. Demuth DR, Golub EE, Malamud D. 1990. Streptococcal-host interactions. Structural and functional analysis of a Streptococcus sanguis receptor for a human salivary glycoprotein. J. Biol. Chem. 265:7120-26

41. Demuth DR, Lammey MS, Huck M, Lally ET, Malamud D. 1990. Comparison of Streptococcus mutans and Streptococcus sanguis receptors for human salivary agglutinin. Microb. Pathog. 9:199-211

42. Deleted in proof
43. Donkersloot JA, Cisar JO, Wax ME, Harr RJ, Chassy BM. 1985. Expression of Actinomyces viscosus antigens in Escherichia coli: cloning of a structural gene (fimA) for type 2 fimbriae. J. Bacteriol. 162:1075-78
44. Duan Y, Fisher E, Malamud D, Golub E, Demuth DR. 1994. Calcium-binding properties of SSP-5, the Streptococcus gordonii M5 receptor for salivary agglutinin. Infect. Immun. 62:5220-26
45. Duchesne P, Grenier D, Mayrand D. 1995. Demonstration of adherence properties of Porphyromonas gingivalis outer membrane vesicles using a new microassay. Oral Microbiol. Immunol. 10:76-80
46. Duncan MJ, Nakao S, Skobe Z, Xie H. 1993. Interactions of Porphyromonas gingivalis with epithelial cells. Infect. Immun. 61:2260-65
47. Ebisu S, Nakae H, Okada H. 1988. Coaggregation of Eikenella corrodens with oral bacteria mediated by bacterial lectin-like substance. Adv. Dent. Res. 2:323-27
48. Egli C, Leung WK, Muller KH, Hancock REW, McBride BC. 1993. Poreforming properties of the major 53-kilodalton surface antigen from outer sheath of Treponema denticola. Infect. Immun. 61:1694-99
49. Ellen RP, Dawson JR, Yang PF. 1994. Treponema denticola as a model for polar adhesion and cytopathogenicity of spirochetes. Trends Microbiol. 2:114-19
50. Ellen RP, Veisman H, Buivids IA, Rosenberg M. 1994. Kinetics of lactosereversible coadhesion of Actinomyces naeslundii WVU398A and Streptococcus oralis 34 on the surface of hexadecane droplets. Oral Microbiol. Immunol. 9:364-71
51. Engelman DM, Steitz TA, Goldman A. 1986. Identifying nonpolar transbilayer helices in amino acid sequences of membrane proteins. Annu. Rev. Biophys. Biophys. Chem. 15:321-53
52. Fenno JC, LeBlanc DJ, Fives-Taylor P. 1989. Nucleotide sequence analysis of a type 1 fimbrial gene of Streptococcus sanguis FW213. Infect. Immun. 57:3527-33
53. Fenno JC, Shaikh A, Fives-Taylor P. 1993. Characterization of allelic replacement in Streptococcus parasanguis: transformation and homologous recombination in a "nontransformable" streptococcus. Gene 130:81-90
54. Fenno JC, Shaikh A, Spatafora G, Fives-Taylor P. 1995. The fimA locus of Streptococcus parasanguis encodes an ATP-binding membrane transport system. Mol. Microbiol. 15:849-63
55. Firth N, Fink PD, Johnson L, Skurray RA. 1994. A lipoprotein signal peptide encoded by the staphylococcal conjugative plasmid pSK41 exhibits an activity resembling that of Enterococcus faecalis pheromone cAD1. J. Bacteriol. 176:5871-73
56. Fives-Taylor P, Meyer D, Mintz K. 1995. Characteristics of Actinobacillus actinomycetemcomitans invasion of and adhesion to cultured epithelial cells. Adv. Dent. Res. 9:55-62
57. Fives-Taylor PM, Meyer DH, Sreenivasan PK, Mintz KP. 1994. Invasion of cultured epithelial cells by periodontopathogens. In Molecular Pathogenesis of Periodontal Disease, ed. R Genco, S Hamada, T Lehner, JR McGhee, S Mergenhagen, pp. 57-68. Washington, DC: Am. Soc. Microbiol.
58. Fuqua WC, Winans SC, Greenberg EP. 1994. Quorum sensing in bacteria: the LuxR-LuxI family of cell density-responsive transcriptional regulators. J. Bacteriol. 176:269-75
59. Ganeshkumar N, Arora N, Kolenbrander PE. 1993. Saliva-binding protein (SsaB) from Streptococcus sanguis 12 is a lipoprotein. J. Bacteriol. 175:572-74
60. Ganeshkumar N, Hannam PM, Kolenbrander PE, McBride BC. 1991. Nucleotide sequence of a gene coding for a saliva-binding protein (SsaB) from Streptococcus sanguis 12 and possible role of the protein in coaggregation with actinomyces. Infect. Immun. 59:1093-99
61. Deleted in proof
62. George KS, Falkler WA Jr. 1992. Coaggregation studies of the Eubacterium species. Oral Microbiol. Immunol. 7:285-90
63. Gibbons RJ, Hay DI. 1988. Adsorbed salivary proline-rich proteins as bacterial receptors on apatitic surfaces. In Molecular mechanisms of microbial adhesion, ed. L Switalski, M Hook, E Beachey, pp. 143-63. New York: Springer
64. Gibbons RJ, Hay DI, Cisar JO, Clark WB. 1988. Adsorbed salivary proline-rich protein 1 and statherin: receptors for type 1 fimbriae of Actinomyces viscosus T14V-J1 on apatitic surfaces. Infect. Immun.

56:2990-93
65. Gibbons RJ, Hay DI, Schlesinger DH. 1991. Delineation of a segment of adsorbed salivary acidic proline-rich proteins which promotes adhesion of Streptococcus gordonii to apatitic surfaces. Infect. Immun. 59:2948-54
66. Gibbons RJ, Nygaard M. 1970. Interbacterial aggregation of plaque bacteria. Arch. Oral. Biol. 15:1397-1400
67. Gillece-Castro BL, Prakobphol A, Burlingame AL, Leffler H, Fisher SJ. 1991. Structure and bacterial receptor activity of a human salivary proline-rich glycoprotein. J. Biol. Chem. 266:17358-68
68. Gilson E, Alloing G, Schmidt T, Claverys J-P, Dudler R, Hofnung M. 1988. Evidence for high affinity binding-protein dependent transport systems in grampositive bacteria and inMycoplasma. EMBO J. 7:3971-74
69. Glushka J, Cassels FJ, Carlson RW, van Halbeek H. 1992. Complete structure of the adhesin receptor polysaccharide of Streptococcus oralis ATCC 55229 (Streptococcus sanguis H1). Biochemistry 31:10741-46
70. Goulbourne PA, Ellen RP. 1991. Evidence that Porphyromonas (Bacteroides) gingivalis fimbriae function in adhesion to Actinomyces viscosus. J. Bacteriol. 173:5266-74
71. Grenier D. 1992. Demonstration of a bimodal coaggregation reaction between Porphyromonas gingivalis and Treponema denticola. Oral Microbiol. Immunol. 7:280-84
72. Grenier D. 1992. Nutritional interactions between two suspected periodontopathogens, Treponema denticola and Porphyromonas gingivalis. Infect. Immun. 60:5298-5301
73. Grenier D, Mayrand D. 1987. Functional characterization of extracellular vesicles produced by Bacteroides gingivalis. Infect. Immun. 55:111-17
74. Haapasalo M, Muller KH, Uitto V-J, Leung WK, McBride BC. 1992. Characterization, cloning, and binding properties of the major 53-kilodalton Treponema denticola surface antigen. Infect. Immun. 60:2058-65
75. Handley PS. 1990. Structure, composition, and functions of surface structures on oral bacteria. Biofouling 2:239-64
76. Hasty DL, Ofek I, Courtney HS, Doyle RJ. 1992. Multiple adhesins of streptococci. Infect. Immun. 60:2147-52
77. Hayashi S, Wu HC. 1990. Lipoproteins in bacteria. J. Bioenerg. Biomembr. 22:451-71
78. Hayashi S, Wu HC. 1992. Identification and characterization of lipid-modified proteins in bacteria. In Lipid Modification of Proteins: A Practical Approach, ed. NM Hooper, AG Turner, pp. 261-68. Oxford: Oxford Univ. Press
79. Higgins CF. 1992. **ABC transporters** : from microorganisms to man. Annu. Rev. Cell Biol. 8:67-113
80. Higgins CF, Gallagher MP, Mimmack ML, Pearce SR. 1988. A family of closely related ATP-binding subunits from prokaryotic and eukaryotic cells. Bioessays 8:111-16
81. Hsu SD, Cisar JO, Sandberg AL, Kilian M. 1994. Adhesive properties of viridans streptococcal species. Microb. Ecol. Health Dis. 7:125-37
82. Hughes CV, Andersen RN, Kolenbrander PE. 1992. Characterization of Veillonella atypica PK1910 adhesin-mediated coaggregation with oral Streptococcus spp. Infect. Immun. 60:1178-86
83. Hughes CV, Roseberry CA, Kolenbrander PE. 1990. Isolation and characterization of coaggregation-defective mutants of Veillonella atypica. Arch. Oral Biol. 35:123S-25S
84. Hyde SC, Emsley P, Hartshorn MJ, Mimmack MM, Gileadi U, et al. 1990. Structural model of ATP-binding proteins associated with cystic fibrosis, multidrug resistance and bacterial transport. Nature 346:362-65
85. Isogai E, Hirose K, Fujii N, Isogai H. 1992. Three types of binding by Porphyromonas gingivalis and oral bacteria to fibronectin, buccal epithelial cells, and erythrocytes. Arch. Oral Biol. 37:667-70
86. Jenkinson HF. 1992. Adherence, coaggregation, and hydrophobicity of Streptococcus gordonii associated with expression of cell surface lipoproteins. Infect. Immun. 60:1225-28
87. Jenkinson HF. 1994. Adherence and accumulation of oral streptococci. Trends Microbiol. 2:209-12
88. Jenkinson HF. 1994. Cell surface protein receptors in oral streptococci. FEMS Microbiol. Lett. 121:133-40

89. Jenkinson HF, Baker RA, Tannock GW. 1996. A binding-lipoprotein dependent oligopeptide transport system in Streptococcus gordonii essential for uptake of hexa-and heptapeptides. J. Bacteriol. 178:68-77

90. Jenkinson HF, Easingwood RA. 1990. Insertional inactivation of the gene encoding a 76-kilodalton cell surface polypeptide in Streptococcus gordonii Challis has a pleiotropic effect on cell surface composition and properties. Infect. Immun. 58:3689-97

91. Jenkinson HF, Terry SD, McNab R, Tannock GW. 1993. Inactivation of the gene encoding surface protein SspA in Streptococcus gordonii DL1 affects cell interactions with human salivary agglutinin and oral actinomyces. Infect. Immun. 61:3199-3208

92. Johnson JL, Moore LVH, Kaneko B, Moore WEC. 1990. Actinomyces georgiae sp. nov., Actinomyces gerencseriae sp. nov., designation of two genospecies of Actinomyces naeslundii, and inclusion of A. naeslundii serotypes II and III and Actinomyces viscosus serotype II in A. naeslundii genospecies 2. Int. J. Syst. Bacteriol. 40:273-86

93. Kagermeier A, London J. 1986. Identification and preliminary characterization of a lectin-like protein from Capnocytophaga gingivalis (emended). Infect. Immun. 51:490-94

94. Kaufman J, DiRienzo JM. 1989. Isolation of a corncob (coaggregation) receptor polypeptide from Fusobacterium nucleatum. Infect. Immun. 57:331-37

95. Kelly CG, Todryk S, Kendal HL, Munro GH, Lehner T. 1995. T-cell, adhesion, and B-cell epitopes of the cell surface Streptococcus mutans protein antigen I/II. Infect. Immun. 63:3649-58

96. Kigure T, Saito A, Seida K, Yamada S, Ishihara K, Okuda K. 1995. Distribution of Porphyromonas gingivalis and Treponema denticola in human subgingival plaque at different periodontal pocket depths examined by immunohistochemical methods. J. Periodont. Res. 30:332-41

97. Kinder S, Holt SC. 1989. Characterization of coaggregation between Bacteroides gingivalis T22 and Fusobacterium nucleatum T18. Infect. Immun. 57:3425-33

98. Kinder SA, Holt SC. 1993. Localization of the Fusobacterium nucleatum T18 adhesin activity mediating coaggregation with Porphyromonas gingivalis T22. J. Bacteriol. 175:840-50

99. Deleted in proof

100. Klier CM, Roble AG, Kolenbrander PE, Marco ML, Cross S, Handley PS. 1996. Identification of a 95-kDa putative adhesin from Actinomyces serovar WVA963 that binds to coaggregation partner Streptococcus oralis 34. Microbiology. Submitted

101. Koga T, Okahashi N, Takahashi I, Kanamoto T, Asakawa H, Iwaki M. 1990. Surface hydrophobicity, adherence, and aggregation of cell surface protein antigen mutants of Streptococcus mutans serotype c. Infect, Immun. 58:289-96

102. Kolenbrander PE. 1988. Intergeneric coaggregation among human oral bacteria and ecology of dental plaque. Annu. Rev. Microbiol. 42:627-56

103. Kolenbrander PE. 1989. Surface recognition among oral bacteria: multigeneric coaggregations and their mediators. Crit. Rev. Microbiol. 17:137-59

104. Kolenbrander PE. 1991. Coaggregation: adherence in the human oral microbial ecosystem. In Microbial Cell-Cell Interactions, ed. M Dworkin, p. 303-29. Washington, DC: Am. Soc. Microbiol.

105. Kolenbrander PE. 1995. Coaggregations among oral bacteria. In Adhesion of Microbial Pathogens, ed. RJ Doyle, I Ofek, p. 385-97. Methods in Enzymology, vol. 253., San Diego: Academic

106. Kolenbrander PE, Andersen RN. 1984. Cell-to-cell interactions of Capnocytophaga and Bacteroides species with other oral bacteria and their potential role in development of plaque. J. Periodontal Res. 19:564-69

107. Kolenbrander PE, Andersen RN. 1986. Multigeneric aggregations among oral bacteria: a network of independent cell-to-cell interactions. J. Bacteriol. 168:851-59

108. Kolenbrander PE, Andersen RN. 1989. Inhibition of coaggregation between Fusobacterium nucleatum and Porphyromonas (Bacteroides) gingivalis by lactose and related sugars. Infect. Immun. 57:3204-9

109. Kolenbrander PE, Andersen RN. 1990. Characterization of Streptococcus gordonii (S. sanguis) PK488 adhesinmediated coaggregation with Actinomyces naeslundii PK606. Infect. Immun. 58:3064-72

110. Kolenbrander PE, Andersen RN, Ganeshkumar N. 1994. Nucleotide sequence of the Streptococcus gordonii PK488 coaggregation adhesin gene, scaA, and ATP-binding cassette. Infect. Immun. 62:4469-80

111. Kolenbrander PE, Andersen RN, Holdeman LV. 1985. Coaggregation of oral Bacteroides species with other bacteria: central role in coaggregation bridges and competitions. Infect. Immun. 48:741-46

112. Kolenbrander PE, Andersen RN, Moore LVH. 1989. Coaggregation of Fusobacterium nucleatum, Selenomonas flueggei, Selenomonas infelix, Selenomonas noxia, and Selenomonas sputigena with strains from 11 genera of oral bacteria. Infect. Immun. 57:3194-3203

113. Kolenbrander PE, Andersen RN, Moore LVH. 1990. Intrageneric coaggregation among strains of human oral bacteria: potential role in primary colonization of the tooth surface. Appl. Environ. Microbiol. 56:3890-94

114. Kolenbrander PE, Celesk RA. 1983. Coaggregation of human oral Cytophaga species and Actinomyces israelii. Infect. Immun. 40:1178-85

115. Kolenbrander PE, Ganeshkumar N, Cassels FJ, Hughes CV. 1993. Coaggregation: specific adherence among human oral plaque bacteria. FASEB J. 7:406-13

116. Deleted in proof

117. Kolenbrander PE, Inouye Y, Holdeman LV. 1983. New Actinomyces and Streptococcus coaggregation groups among human oral isolates from the same site. Infect. Immun. 41:501-6

118. Kolenbrander PE, London J. 1992. Ecological significance of coaggregation among oral bacteria. Adv. Microb. Ecol. 12:183-217

119. Kolenbrander PE, London J. 1993. Adhere today, here tomorrow: oral bacterial adherence. J. Bacteriol. 175:3247-52

120. Kolenbrander PE, Parrish KD, Andersen RN, Greenberg EP. 1995. Intergeneric coaggregation of oral Treponema spp. with Fusobacterium spp. and intrageneric coaggregation among Fusobacterium spp. Infect. Immun. 63:4584-88

121. Lai C-H, Bloomquist C, Liljemark WF. 1990. Purification and characterization of an outer membrane protein adhesin from Haemophilus parainfluenzae HP-28. Infect. Immun. 58:3833-39

122. Lamont RJ, Bevan CA, Gil S, Persson RE, Rosan B. 1993. Involvement of Porphyromonas gingivalis fimbriae in adherence to Streptococcus gordonii. Oral Microbiol. Immunol. 8:272-76

123. Lamont RJ, Chan A, Belton CM, Izutsu KT, Vasel D, Weinberg A. 1995. Porphyromonas gingivalis invasion of gingival epithelial cells. Infect. Immun. 63:3878-85

124. Lamont RJ, Demuth DR, Davis CA, Malamud D, Rosan B. 1991. Salivaryagglutinin--mediated adherence of Streptococcus mutans to early plaque bacteria. Infect. Immun. 59:3446-50

125. Lamont RJ, Gil S, Demuth DR, Malamud D, Rosan B. 1994. Molecules of Streptococcus gordonii that bind to Porphyromonas gingivalis. Microbiology 140:867-72

126. Lamont RJ, Hersey SG, Rosan B. 1992. Characterization of the adherence of Porphyromonas gingivalis to oral streptococci. Oral Microbiol. Immunol. 7:193-97

127. Lamont RJ, Oda D, Persson RE, Persson GE. 1992. Interaction of Porphyromonas gingivalis with gingival epithelial cells maintained in culture. Oral Microbiol. Immunol. 7:364-67

128. Lamont RJ, Rosan B. 1989. Characteristics of a protease of Streptococcus sanguis G9B which degrades the major salivary adhesin. FEMS Microbiol. Lett. 58:101-6

129. Lamont RJ, Rosan B, Baker CT, Nelson GM. 1988. Characterization of an adhesin antigen of Streptococcus sanguis G9B. Infect. Immun. 56:2417-2423

130. Lantz MS, Allen RD, Duck LW, Vail TA, Blume JL, et al. 1991. Identification of components of Porphyromonas gingivalis that mediate its interactions with fibronectin. J. Bacteriol. 173:4263-70

131. Lantz MS, Allen RD, Vail TA, Switalski LM, Hook M. 1991. Specific cell components of Bacteroides gingivalis mediate binding and degradation of human fibrinogen. J. Bacteriol. 173:495-504

132. LaPolla RJ, Haron JA, Kelly CG, Taylor WR, Bohart C, et al. 1991. Sequence and structural analysis of surface protein antigen I/II (SpaA) of Streptococcus sobrinus. Infect. Immun. 59:2677-85

133. Lee J-Y, Sojar HT, Bedi GS, Genco RJ. 1992. Synthetic peptides analogous to the fimbrillin sequence inhibit adherence of Porphyromonas

gingivalis. Infect. Immun. 60:1661-70
134. Leonhardt A, Olsson J, Dahlen G. 1995. Bacterial colonization on titanium, hydroxyapatite, and amalgam surfaces in vivo. J. Dent. Res. 74:1607-12
135. Li J, Ellen RP, Hoover CI, Felton JR. 1991. Association of proteases of Porphyromonas (Bacteroides) gingivalis with its adhesion to Actinomyces viscosus. J. Dent. Res. 70:82-86
136. Ligtenberg AJM, Walgreen-Weterings E, Veerman ECI, de Soet JJ, de Graaff J, Nieuw Amerongen AV. 1992. Influence of saliva on aggregation and adherence of Streptococcus gordonii HG 222. Infect. Immun. 60:3878-84
137. Liljemark WF, Bloomquist CG, Fenner LJ. 1985. Characteristics of the adherence of oral Haemophilus species to an experimental salivary pellicle and to other oral bacteria. In Molecular Basis of Oral Microbial Adhesion, ed. SE Mergenhagen, B Rosan, pp. 94-102. Washington, DC: Am. Soc. Microbiol.
138. Lo CS, Hughes CV. 1996. Identification and characterization of a protease from Streptococcus oralis C104. Oral Microbiol. Immunol. 11:181-85
139. Loach DM, Jenkinson HF, Tannock GW. 1994. Colonization of the murine oral cavity by Streptococcus gordonii. Infect. Immun. 62:2129-31
140. London J, Allen J. 1990. Purification and characterization of a Bacteroides loeschei adhesin that interacts with procaryotic and eucaryotic cells. J. Bacteriol. 172:2527-34
141. London J, Hand AR, Weiss AI, Allen J. 1989. Bacteroides loeschei PK 1295 cells express two distinct adhesins simultaneously. Infect. Immun. 57:3940-44
142. Lowe AM, Lambert PA, Smith AW. 1995. Cloning of an Enterococcus faecalis endocarditis antigen: homology with adhesins from some oral streptococci. Infect. Immun. 63:703-6
143. Lunsford RD. 1995. A Tn4001 delivery system for Streptococcus gordonii (Challis). Plasmid 33:153-57
144. Ma JK-C, Kelly CG, Munro G, Whiley RA, Lehner T. 1991. Conservation of the gene encoding streptococcal antigen I/II in oral streptococci. Infect. Immun. 59:2686-94
145. Malek R, Fisher JG, Caleca A, Stinson M, van Oss CJ, et alGenco RJ. 1994. Inactivation of the Porphyromonas gingivalis fimA gene blocks periodontal damage in gnotobiotic rats. J. Bacteriol. 176:1052-59
146. Manch-Citron JN, Allen J, Moos JM, London J. 1992. The gene encoding a Prevotella loescheii lectin-like adhesin contains an interrupted sequence which causes a frameshift. J. Bacteriol. 174:7328-36
147. Manch-Citron JN, London J. 1994. Expression of the Prevotella adhesin gene (plaA) is mediated by a programmed frameshifting hop. J. Bacteriol. 176:1944-48
148. McNab R, Jenkinson HF. 1992. Gene disruption identifies a 290-kDa cellsurface polypeptide conferring hydrophobicity and coaggregation properties in Streptococcus gordonii. Mol. Microbiol. 6:2939-49
149. McNab R, Jenkinson HF, Loach DM, Tannock GW. 1994. Cell-surface--associated polypeptides CshA and CshB of high molecular mass are colonization determinants in the oral bacterium Streptococcus gordonii. Mol. Microbiol. 14:743-54
150. Meyer DH, Fives-Taylor PM. 1993. Evidence that extracellular components function in adherence of Actinobacillus actinomycetemcomitans to epithelial cells. Infect. Immun. 61:4933-36
151. Meyer DH, Fives-Taylor PM. 1994. Characteristics of adherence of Actinobacillus actinomycetemcomitans to epithelial cells. Infect. Immun. 62:928-35
152. Meyer DH, Sreenivasan PK, Fives-Taylor PM. 1991. Evidence for invasion of a human oral cell line by Actinobacillus actinomycetemcomitans. Infect. Immun. 59:2719-26
153. Moore WEC, Moore LVH. 1994. The bacteria of periodontal diseases. Periodontol. 2000 5:66-77
154. Mouton C, Reynolds HS, Gasiecki EA, Genco RJ. 1979. In vitro adhesion of tufted oral streptococci to Bacterionema matruchotti. Curr. Microbiol. 3:181-86
155. Murikami Y, Nagata H, Amano A, Takagaki M, Shizukuishi S, et al. 1991. Inhibitory effects of human salivary histatins and lysozyme on coaggregation between Porphyromonas gingivalis and Streptococcus mitis. Infect. Immun. 59:3284-86
156. Murray PA, Prakobphol A, Lee T, Hoover CI, Fisher SJ. 1992.

Adherence of oral streptococci to salivary glycoproteins. Infect. Immun. 60:31-38

157. Nagata H, Amano A, Hanioka T, Tamagawa H, Shizukuishi S, Miyata T. f1993. Inhibition of coaggregation between Porphyromonas gingivalis and Streptococcus oralis by fibrinogen fragments. FEMSMicrobiol. Lett. 114:31-36

158. Nagata H, Amano A, Ojima M, Tanaka M, Kataoka K, Shizukuishi S. 1994. Effect of binding of fibrinogen to each bacterium on coaggregation between Porphyromonas gingivalis and Streptococcus oralis. Oral Microbiol. Immunol. 9:359-63

159. Naito Y, Gibbons RJ. 1988. Attachment of Bacteroides gingivalis to collagenous substrata. J. Dent. Res. 67:1075-80

160. Naito Y, Tohda H, Okuda K, Takazoe I. 1993. Adherence and hydrophobicity of invasive and noninvasive strains of Porphyromonas gingivalis. Oral Microbiol. Immunol. 8:195-202

161. Navarre WW, Schneewind O. 1994. Proteolytic cleavage and cell wall anchoring at the LPXTG motif of surface proteins in gram-positive bacteria. Mol. Microbiol. 14:115-121

162. Nesbitt WE, Beem JE, Leung K-P, Clark WB. 1992. Isolation and characterization of Actinomyces viscosus mutants defective in binding salivary proline-rich proteins. Infect. Immun. 60:1095-1100

163. Nesbitt WE, Fukushima H, Leung K-P, Clark WB. 1993. Coaggregation of Prevotella intermedia with oral Actinomyces species. Infect. Immun. 61:2011-14

164. Nikaido H, Saier MH Jr. 1992. Transport proteins in bacteria: common themes in their design. Science 258:936-42

165. Nyvad B, Kilian M. 1990. Comparison of the initial streptococcal microflora on dental enamel in caries-active and in caries-inactive individuals. Caries Res. 24:267-72

166. Ogawa T, Hamada S. 1994. Hemagglutinating and chemotactic properties of synthetic peptide segments of fimbrial protein from Porphyromonas gingivalis. Infect. Immun. 62:3305-10

167. Oligino L, Fives-Taylor P. 1993. Overexpression and purification of a fibriaassociated adhesin of Streptococcus parasanguis. Infect. Immun. 61:1016-22

168. Pearce BJ, Naughton AM, Masure HR. 1994. Peptide permeases modulate transformation in Streptococcus pneumoniae. Mol. Microbiol. 12:881-92

169. Petsios A, Nakou M, Manti F. 1995. Microflora in adult periodontitis. J. Periodont. Res. 30:325-31

170. Reddy GP, Abeygunawardana C, Bush CA, Cisar JO. 1994. The cell wall polysaccharide of Streptococcus gordonii 38: structure and immunochemical comparison with the receptor polysaccharides of Streptococcus oralis 34 and Streptococcus mitis J22. Glycobiology 4:183-92

171. Reddy GP, Chang C-C, Bush CA. 1993. Determination by heteronuclear NMR spectroscopy of the complete structure of the cell wall polysaccharide of Streptococcus sanguis strain K103. Anal. Chem. 65:913-21

172. Rosan B, Baker CT, Nelson GM, Berman R, Lamont RJ, Demuth DR. 1989. Cloning and expression of an adhesin antigen of Streptococcus sanguis G9B in Escherichia coli. J. Gen. Microbiol. 135:531-38

173. Rosenberg M, Buivids IA, Ellen RP. 1991. Adhesion of Actinomyces viscosus to Porphyromonas (Bacteroides) gingivaliscoated hexadecane droplets. J. Bacteriol. 173:2581-89

174. Ruhfel RE, Manias DA, Dunny GM. 1993. Cloning and characterization of a region of the Enterococcus faecalis conjugative plasmid, pCF10, encoding a sex pheromone-binding function. J. Bacteriol. 175:5253-59

175. Russell H, Tharpe JA, Wells DE, White EH, Johnson JE. 1990. **Monoclonal antibody** recognizing a species-specific protein from Streptococcus pneumoniae. J. Clin. Microbiol. 28:2191-95

176. Russell RRB, Aduse-Opoku J, Sutcliffe IC, Tao L, Ferretti JJ. 1992. A binding protein-dependent transport system in Streptococcus mutans responsible for multiple sugar metabolism. J. Biol. Chem. 267:4631-37

177. Sampson JS, O'Connor SP, Stinson AR, Tharpe JA, Russell H. 1994. Cloning and nucleotide sequence analysis of psaA, the Streptococcus pneumoniae gene encoding a 37-kilodalton protein homologous to previously reported Streptococcus sp. adhesins. Infect. Immun. 62:319-24

178. Sandberg AL, Ruhl S, Joralman RA, Brennan MJ, Sutphin MJ, Cisar JO. 1995. Putative glycoprotein and glycolipid polymorphonuclear leukocyte receptors for the Actinomyces naeslundii WVU45 fimbrial lectin. Infect. Immun. 63:2625-31

179. Sandros J, Papapanou PN, Dahlen G. 1993. Porphyromonas gingivalis invades oral epithelial cells in vitro. J. Periodont. Res. 28:219-26

180. Sandros J, Papapanou PN, Nannmark U, Dahlen G. 1994. Porphyromonas gingivalis invades human pocket epithelium in vitro. J. Periodont. Res. 29:62-69

181. Scannapieco FA. 1994. Saliva-bacterium interactions in oral microbial ecology. Crit. Rev. Oral Biol. Med. 5:203-48

182. Scannapieco FA, Bergey EJ, Reddy MS, Levine MJ. 1989. Characterization of salivary a-amylase binding to Streptococcus sanguis. Infect. Immun. 57:2853-63

183. Scannapieco FA, Torres GI, Levine MJ. 1995. Salivary amylase promotes adhesion of oral streptococci to hydroxyapatite. J. Dent. Res. 74:1360-66

184. Schneewind O, Fowler A, Faull KF. 1995. Structure of the cell wall anchor of surface proteins in Staphylococcus aureus. Science 268:103-6

185. Schneewind O, Model P, Fischetti VA. 1992. Sorting of protein A to the staphylococcal cell wall. Cell 70:267-81

186. Sharma A, Sojar HT, Lee JY, Genco RJ. 1993. Expression of a functional Porphyromonas gingivalis fimbrillin polypeptide in Escherichia coli: purification, physicochemical and immunochemical characterization, and binding characteristics. Infect. Immun. 61:3570-73

187. Shilling KM, Bowen WH. 1992. Glucans synthesized in situ in experimental salivary pellicle function as specific binding sites for Streptococcus mutans. Infect. Immun. 60:284-95

188. Simonson LG, McMahon KT, Childers DW, Morton HE. 1992. Bacterial synergy of Treponema denticola and Porphyromonas gingivalis in a multinational population. Oral Microbiol. Immunol. 7:111-12

189. Singh U, Grenier D, McBride BC. 1989. Bacteroides gingivalis vesicles mediate attachment of streptococci to serumcoated hydroxyapatite. Oral Microbiol. Immunol. 4:199-203

190. Skopek RJ, Liljemark WF. 1994. The influence of saliva on interbacterial adherence. Oral Microbiol. Immunol. 9:19-24

191. Skopek RJ, Liljemark WF, Bloomquist CG, Rudney JD. 1993. Dental plaque development on defined streptococcal surfaces. Oral Microbiol. Immunol. 8:16-23

192. Sreenivasan PK, Meyer DH, Fives-Taylor PM. 1993. Requirements for invasion of epithelial cells by Actinobacillus actinomycetemcomitans. Infect. Immun. 61:1239-45

193. Stinson MW, Haraszthy GG, Zhang SL, Levine MJ. 1992. Inhibition of Porphyromonas gingivalis adhesion to Streptococcus gordonii by human submandibular-sublingual saliva. Infect. Immun. 60:2598-2604

194. Stinson MW, Safulko K, Levine MJ. 1991. Adherence of Porphyromonas (Bacteroides) gingivalis to Streptococcus sanguis in vitro. Infect. Immun. 59:102-8

195. Stromberg N, Boren T. 1992. Actinomyces tissue specificity may depend on differences in receptor specificity for GalNAcb-containing glycoconjugates. Infect. Immun. 60:3268-77

196. Stromberg N, Boren T, Carlen A, Olsson J. 1992. Salivary receptors for GalNAcb-sensitive adherence of Actinomyces spp.: evidence for heterogenous GalNAcb and proline-rich protein receptor properties. Infect. Immun. 60:3278-86

197. Sutcliffe IC, Russell RRB. 1995. Lipoproteins of gram-positive bacteria. J. Bacteriol. 177:1123-28

198. Sutcliffe ID, Tao L, Ferretti JJ, Russell RRB. 1993. MsmE, a lipoprotein involved in sugar transport in Streptococcus mutans. J. Bacteriol. 175:1853-55

199. Takemoto T, Ozaki M, Shirakawa M, Hino T, Okamoto H. 1993. Purification of arginine-sensitive hemagglutinin from Fusobacterium nucleatum and its role in coaggregation. J. Periodont. Res. 28:21-26

200. Tanimoto K, An FY, Clewell DB. 1993. Characterization of the traC determinant of the Enterococcus faecalis hemolysinbacteriocin plasmid pAD1: binding of sex pheromone. J. Bacteriol. 175:5260-64

201. Tempro P, Cassels F, Siraganian R, Hand AR, London J. 1989. Use of adhesinspecific **monoclonal antibodies** to identify and localize an

adhesin on the surface of Capnocytophaga gingivalis DR2001. Infect. Immun. 57:3418-24

202. ter Steeg PF, van der Hoeven JS. 1989. Development of periodontal microflora on human serum. Microb. Ecol. Health Dis. 2:1-10

203. Tokuda M, Okahashi N, Takahashi I, Nakai M, Nagaoka S, et al. 1991. Complete nucleotide sequence of the gene for a surface protein antigen of Streptococcus sobrinus. Infect. Immun. 59:3309-12

204. Tuttle RS, Mangan DF. 1990. Interaction of Fusobacterium nucleatum 191 with human peripheral blood lymphocytes. J. Periodont. Res. 25:364-71

205. Uitto V-J, Pan YM, Leung WK, Larjava H, Ellen RP, et al. 1995. Cytopathic effects of Treponema denticola chymotrypsinlike proteinase on migrating and stratified epithelial cells. Infect. Immun. 63:3401-10

206. Vacca-Smith AM, Jones CA, Levine MJ, Stinson MW. 1994. Glucosyltransferase mediates adhesion of Streptococcus gordonii to human endothelial cells in vitro. Infect. Immun. 62:2187-94

207. Vickerman MM, Sulavik MC, Clewell DB. 1995. Oral streptococci with genetic determinants similar to the glycosyltransferase regulatory gene, rgg. Infect. Immun. 63:4524-27

208. Weerkamp AH, Handley PS, Baars A, Slot JW. 1986. Negative staining and immunoelectron microscopy of adhesiondeficient mutants of Streptococcus salivarius reveal that the adhesive protein antigens are separate classes of cell surface fibril. J. Bacteriol. 165:746-55

209. Weiger R, Netuschil L, van Ohle C, Schlagenhauf U, Brecx M. 1995. Microbial generation time during the early phases of supragingival dental plaque formation. Oral Microbiol. Immunol. 10:93-97

210. Weinberg A, Holt SC. 1991. Chemical and biological activities of a 64-kilodalton outer sheath protein from Treponema denticola strains. J. Bacteriol. 173:6935-6947

211. Weiss EI, Eli I, Shenitzki B, Smorodinsky N. 1990. Identification of the rhamnosesensitive adhesin of Capnocytophaga ochracea ATCC 33596. Arch. Oral Biol. 35:127s-30s

212. Weiss EI, London J, Kolenbrander PE, Andersen RN. 1989. Fimbria-associated adhesin of Bacteroides loeschei taht recognizes receptors on prokaryotic and eukaryotic cells. Infect. Immun. 57:2912-13

213. Weiss EI, London J, Kolenbrander PE, Andersen RN, Fischler C, Siraganian RP. 1988. Characterization of **monoclonal antibodies** to fimbria-associated adhesins of Bacteroides loescheii PK 1295. Infect. Immun. 56:219-24

214. Weiss EI, London J, Kolenbrander PE, Hand AR, Siraganian R. 1988. Localization and enumeration of fimbriaassociated adhesins of Bacteroides loescheii. J. Bacteriol. 170:1123-28

215. Whittaker CJ, Clemans DL, Kolenbrander PE. 1996. Insertional inactivation of an intrageneric coaggregation-relevant adhesin locus from Streptococcus gordonii DL1 (Challis). Infect. Immun. Submitted

216. Willcox MDP, Patrikakis M, Harty DWS, Loo CY, Knox KW. 1993. Coaggregation of oral lactobacilli with streptococci from the oral cavity. Oral Microbiol. Immunol. 8:319-21

217. Wolfaardt GM, Lawrence JR, Robarts RD, Caldwell SJ, Caldwell DE. 1994. Multicellular organization in a degradative biofilm community. Appl. Environ. Microbiol. 60:434-46

218. Yao ES, Lamont RJ, Leu SP, Weinberg A. 1996. Interbacterial binding among strains of pathogenic and commensal oral bacterial species. Oral Microbiol. Immunol. 11:35-41

219. Yeung MK. 1993. Complete nucleotide sequence of the Actinomyces viscosus T14V sialidase gene: presence of a conserved repeating sequence among strains of Actinomyces spp. Infect. Immun. 61:109-16

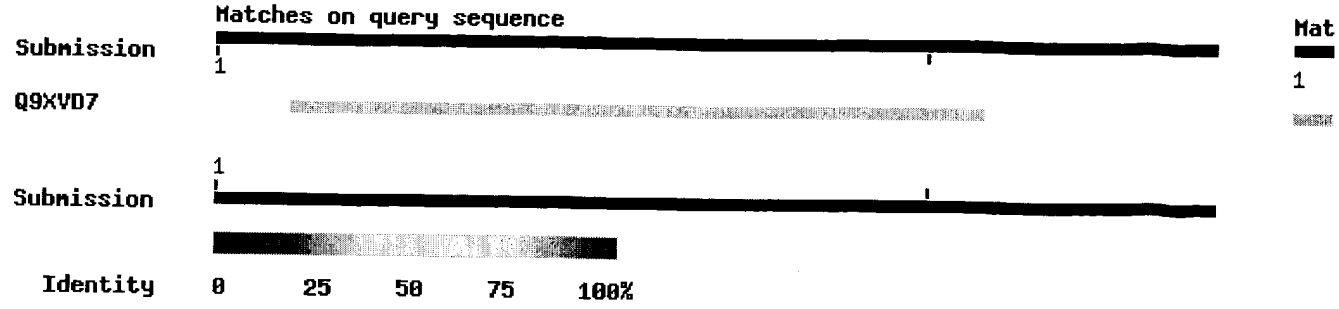
220. Yeung MK. 1995. Construction and use of integration plasmids to generate sitespecific mutations in the Actinomcyes viscosus T14V chromosome. Infect. Immun. 63:2924-30

221. Yeung MK, Chassy BM, Cisar JO. 1987. Cloning and expression of a type 1 fimbrial subunit of Actinomyces viscosus T14V, J. Bacteriol. 169:1678-83

222. Yeung MK, Cisar JO. 1988. Cloning and nucleotide sequence of a gene for Actinomyces naeslundii WVU45 type 2 fimbriae. J. Bacteriol. 170:3803-9

223. Yeung MK, Cisar JO. 1990. Sequence homology between the subunits of two immunologically and functionally distinct types of fimbriae of Actinomyces spp. J. Bacteriol. 172:2462-68

224. Yeung MK, Kozelsky CS. 1994. Transformation of Actinomyces spp. by a gramnegative broad--host-range plasmid. J. Bacteriol. 176:4173-76



Alignments

tr Q9XVD7 **C18D4.9 protein [C18D4.9] [Caenorhabditis elegans]** 361 AA

align

Score = 30.8 bits (65), Expect = 2.3
Identities = 8/10 (80%), Positives = 9/10 (90%)

```
Query: 2   RMYSHRHAHL 11
           RMYS+ HAHL
Sbjct: 244 RMYSQQHAHL 253
```

```
Database: EXPASY/UniProt
    Posted date:  Jun 6, 2004 11:44 AM
  Number of letters in database: 476,696,577
  Number of sequences in database:  1,498,944

Lambda      K       H
   0.346    0.265     2.13

Gapped
Lambda      K       H
   0.294    0.110    0.610

Matrix: PAM30
Gap Penalties: Existence: 9, Extension: 1
Number of HSP's successfully gapped in prelim test: 0
length of query: 14
length of database: 476,696,577
effective HSP length: 5
effective length of query: 9
effective length of database: 469,201,857
effective search space: 4222816713
effective search space used: 4222816713
T: 16
A: 40
X1: 15 ( 7.5 bits)
X2: 35 (14.8 bits)
X3: 58 (24.6 bits)
S1: 40 (21.9 bits)
S2: 61 (29.1 bits)
```

==========

Welcome to the SIB BLAST Network Service

If results of this search are reported or published, please mention that the computation was performed at the SIB using the BLAST network service. The SIB BLAST network service uses a server developed at SIB and the NCBI BLAST 2 software.

In case of problems, please read the online BLAST help.
If your question is not covered, please contact <helpdesk@expasy.org>.

NCBI BLAST program reference [PMID:9254694]:
Altschul S.F., Madden T.L., Schäffer A.A., Zhang J., Zhang Z., Miller W., Lipman D.J. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25:3389-3402(1997).

==========

Query length: 14 AA
Date run: 2004-06-09 22:48:29 UTC+0100 on sib-gm1.unil.ch
Program: NCBI BLASTP 1.5.4-Paracel [2003-06-05]
Database: EXPASY/UniProt
1,498,944 sequences; 476,696,577 total letters

Taxonomic view | NiceBlast view | Printable view

List of potentially matching sequences

Send selected sequences to Clustal W (multiple alignment) Submit Query

Select up to

☐ Include query sequence

	Db AC	Description	Score	E-value
☐	tr Q9XVD7	C18D4.9 protein [C18D4.9] [Caenorhabditis elegans]	31	2.3

Graphical overview of the alignments

Click here to resubmit your query after masking regions matching PROSITE profiles or Pfam HMMs

(? Help) (use ScanProsite for more details about PROSITE matches)

Profile hits

Pfam hits

Genome Sequence of the Nematode *C. elegans*: A Platform for Investigating Biology

The *C. elegans* Sequencing Consortium*

REVIEW

The 97-megabase genomic sequence of the nematode *Caenorhabditis elegans* reveals over 19,000 genes. More than 40 percent of the predicted protein products find significant matches in other organisms. There is a variety of repeated sequences, both local and dispersed. The distinctive distribution of some repeats and highly conserved genes provides evidence for a regional organization of the chromosomes.

The genome sequence of *C. elegans* is essentially complete. The sequence follows those of viruses, several bacteria, and a yeast (*1, 2*) and is the first from a multicellular organism. Some small gaps remain to be closed, but this will be a prolonged process without much biological return. It therefore now makes sense to review the project as a whole.

Here, we describe the origins of the project, the reasons for undertaking it, and the methods that have been used, and we provide a brief overview of the analytical findings. The project began with the development of a clone-based physical map (*3, 4*) to facilitate the molecular analysis of genes, which were being discovered at an ever increasing pace through the study of mutants. This, in turn, initiated a collaboration between the *C. elegans* Sequencing Consortium and the entire community of *C. elegans* researchers (*5*). The resulting free exchange of data and the immediate release of map information (and later sequence) have been hallmarks of the project. The resultant cross correlation between physical and genetic maps is ongoing and is essential for achieving an increasing utility of the sequence.

Along with the genome sequencing project, expressed sequence tag (EST) sequencing has been carried out. Early surveys of expressed sequences were conducted (*6*), but complementary DNA (cDNA) analysis has been carried out primarily by Y. Kohara (*7*). This group has contributed 67,815 ESTs from 40,379 clones, representing an estimated 7432 genes. This extensive information has been invaluable in identifying and annotating genes in the genomic sequence. Others also contributed the 15-kilobase (kb) mitochondrial genome sequence (*8*).

Sequencing

The preexisting physical map, on which sequencing was based, had been initiated by the isolation and assembly of random cosmid clones (with a 40-kb insert, which was the largest insert cloning system available at the time) with a fingerprinting method (*3*). At a sixfold redundant coverage of the genome in cosmids, nonrandom gaps persisted. In most cases, hybridization screening of cosmid libraries failed to yield bridging clones, but the newly developed yeast artificial chromosome (YAC) clones (*9*) rapidly closed most of the cosmid gaps. Incidentally, the YAC clones also covered almost all of the genome, providing a convenient tool for the rapid scanning of the entire genome by hybridization (*4*). About 20% of the genome is represented only in YACs.

By 1989, it became apparent that, with the physical map in hand, complete sequencing of the genome might be both feasible and desirable. Joint funding [from the National Institutes of Health and the UK Medical Research Council (MRC)] for a pilot study was arranged, and in 1990, the first 3-megabase (Mb) sequence was undertaken. Success in this venture (*10, 11*) resulted in full funding and the expansion of the two groups of the consortium in 1993.

Sequencing began in the centers of the chromosomes, where cosmid coverage and the density of genetic markers are high. Cosmids were selected by fingerprint analysis to achieve a tiling path of overlapping clones (in practice, 25% overlap on average). Some sequencing of YACs was explored (*12*), but because of yeast DNA that contaminated preparations of YAC DNA, this approach was deferred in anticipation of the complete sequence of yeast, which enabled contaminating reads to be easily identified. The sequencing process (*13*) can be divided into two major parts: the shotgun phase, which is sequence acquisition from random subclones, and the finishing phase, which is directed sequence acquisition to close any remaining gaps and to resolve ambiguities and low-quality areas. Numerous and ongoing improvements to the shotgun phase have increased sequencing efficiency, improved data quality, and lowered costs. Similarly, finishing tools have improved dramatically. Nonetheless, finishing still requires substantial manual intervention, with a variety of specialized techniques (*14, 15*).

Restriction digests with several enzymes were performed on most cosmids and provided valuable checks on sequence assembly. Where assembly was ambiguous because of repeats, the digests were helpful in resolving the problem. At the start of the project, polymerase chain reaction (PCR) checks were conducted along the length of the sequence to confirm that the assembled sequence of the bacterial clone was an accurate representation of the genome. These checks were abandoned after it became clear that failures in PCR were more common than discrepancies between the clone and the genome.

When available cosmids were exhausted, we screened fosmids (which are similar to cosmids but are maintained at a single copy per cell and thus are potentially more stable) (*16*) and found that a third of the gaps were bridged in the central regions of the chromosomes but very few were bridged in the outer regions. We also used long-range PCR (*17*) to recover some of the central gaps. The remainder of the central gaps and all of the gaps in the outer regions were recovered by sequencing YACs. As for the cosmids, a tiling path of YACs was chosen, and DNA from selected clones was isolated by pulsed-field gel electrophoresis (*18*). Sequencing was performed as for cosmids, with suitable adaptations for the smaller amount of DNA that was available for making libraries. Restriction digests were carried out for assembly checks, but they were not as precisely interpretable as those for bacterial clones. At this stage, the physical map was consolidated and sometimes rearranged as the YAC sequences confirmed or rejected the links made previously by hybridization. The comparison of the assembled YAC sequences with the often extensively overlapping cosmid sequences showed few discrepancies between the two sequences. Generally, further investigation revealed that most discrepancies resulted from a rearrangement in the cosmid. It is interesting (and crucial to the success of the YAC sequencing) that nearly all regions of the YACs can be cloned in

*See genome.wustl.edu/gsc/C_elegans/ and www.sanger.ac.uk/Projects/C_elegans/ for a list of authors. Address correspondence to The Washington University Genome Sequencing Center, Box 8501, 4444 Forest Park Parkway, St. Louis, MO 63108, USA. E-mail: worm@watson.wustl.edu; or The Sanger Centre, The Wellcome Trust Genome Campus, Hinxton, Cambridgeshire CB10 1SA, UK. E-mail: worm@sanger.ac.uk

bacteria as short fragments, although cosmid and fosmid libraries failed to represent these regions.

The key step in closing sequence assemblies was to obtain subclones that bridged the gaps remaining after the shotgun phase. Often, gaps are spanned by the subclones used in the shotgun phase, because the insert length is deliberately set at two to four times the typical sequence read length. The introduction of plasmid clones halfway through the program greatly improved the coverage of inverted repeats and other unusual structures. In cases where the shotgun phase failed to yield a spanning subclone, plasmid clones that bridged gaps were obtained by isolating and subcloning restriction fragments from cosmids. In YACs, because of their greater size and complexity, screening by hybridization was necessary to recover the desired subclone. In the most difficult cases, we have exploited very short insert plasmid libraries to find gap-bridging clones. PCR was used occasionally, but because of its tendency to yield artifacts in repeat regions, it has recently been used as little as possible. Once isolated, the gap-bridging clone was either sequenced directly or, in cases of a difficult secondary structure, a short insert library (SIL) was made by breaking the insert of the gap-bridging clone into smaller fragments (0.5 kb or even smaller in difficult cases), with break points interrupting the secondary structure (*15*). In some cases, transposon insertion has been used (*19*), although SILs are generally preferred as a first pass because of their ease of throughput.

The 97-Mb sequence is a composite of 2527 cosmids, 257 YACs, 113 fosmids, and 44 PCR products (*20, 21*). For the 12 chromosome ends, nine of the telomere plasmid clones provided by Wicky *et al.* have been linked to the outermost YACs (*22*), either directly by sequence or by long-range PCR and sequencing, where no direct sequence link was found. This probably represents >99% of the genome sequence, on the basis of the representation in the genomic sequence of available EST data and of the sequence from random clones from a whole-genome library.

Much of the remaining DNA likely resides in the three residual gaps between the telomeres and the outermost sequenced YACs and in two internal gaps, where no spanning YAC clone has been identified. One of these is known to be <450 kb, on the basis of Southern (DNA) analysis, but a reliable size estimate is not available for the other gaps. A smaller amount will be recovered from four smaller segments (which are spanned by YACs), where shotgun sequencing has not been completed. Furthermore, very small segments (likely to be <1 kb each) have not been recovered in subclones for 139 segments. Finally, some sequence is likely to be missing from the large tandem repeats, which, in extreme cases, consist of tens of kilobases that are composed of hundreds of copies of a short sequence. Although most have been sized by restriction digestion of the cloned DNA, some segments in the larger YACs are of unknown size. Having established the repeat elements, we cannot usefully work further on them at this stage, because they are likely to be variable and because they do not clone stably; any repeat elements that prove to be important will become the subject of population studies in the future.

As shown by the resolution of discrepancies resulting from matches with sequence data from other sources, the error rate of almost all the product is $<10^{-4}$. In a few regions (predominantly in regions of extensive tandem repeats), the sequence is tagged to indicate that a lower standard of accuracy has been accepted. Accuracy is maintained by a set of criteria (*23*), which is followed by the finisher and by a final checking step that requires specialized software (*24*) and a visual inspection. None of this, however, overcomes errors in the cloning process. A comparison of different clones in overlapping regions and the resolution of discrepancies have indicated a finite error rate associated with cloning. For example, cosmid B0393 (GenBank accession number Z37983) contains a deletion of a large hairpin that was only detected because it overlapped cosmid F17C8 (GenBank accession number Z35719); similarly, we detected a 400–base pair region that had been deleted in all M13 and PCR reads from cosmid F59D12 (GenBank accession number Z81558). The F59D12 deletion was detected by restriction digestion and was recovered in plasmids. However, these instances are rare enough that undetected errors are likely to be few; thus, the advantages of the clone-based sequence, in avoiding long-range confusion in assembly, more than make up for its occasional defects.

Sequence Content

Whereas the sequencing has essentially been completed, analysis and annotation will continue for many years, as more information and better sequence annotation tools become available.

To begin the task, we subjected each completed segment to a series of automatic analyses to reveal possible protein (*25*) and transfer RNA (tRNA) genes (*26*), similarities to ESTs and other proteins (*27–30*), repeat families, and local repeats (*31*). The results were entered in the genome database "a *C. elegans* database" (ACEDB) (*32*), which merges overlapping sequences to provide seamless views across clone boundaries and allows the periodic and automatic updating of entries. To integrate and reconcile the various views of the sequence, we reviewed all data interactively through the ACEDB annotator's graphical workbench (*32*). In particular, the GENEFINDER (*25*) predictions are confirmed or adjusted to account for protein, cDNA, and EST matches, repeats, and so forth, and annotation concerning putative gene function is added.

The interruption of the coding sequence by introns, the generation of alternatively spliced forms, and the relatively low gene density make accurate gene prediction more challenging in multicellular organisms than in microbial genomes. The problem is made more complex in *C. elegans* by transplicing and by the organization of as many as 25% of the genes into operons (*33*). We have used GENEFINDER to identify putative coding regions and to provide an initial overview of gene structure. To quantitate the accuracy of gene identification, we compared introns that were confirmed by ESTs and cDNAs to those that were predicted by GENEFINDER. We found that 92% of the predicted introns had an exact match to the experimentally confirmed ones and that 97% had an overlap. Identification of the start and stop of genes is more difficult, and errors in this process sometimes result in the merging of some neighboring genes and in the splitting of others. To refine the computer-generated gene structure predictions, expert annotators use any available EST and protein similarities, as well as genomic sequence data from the related nematode *C. briggsae*. This information can be especially important in establishing gene boundaries. About 40% of the predicted genes have a confirming EST match, but because ESTs are partial, they presently confirm only ~15% of the total coding sequence. In a number of cases, ESTs have provided direct evidence of alternative splicing; these instances have been annotated in the sequence (*34*).

The genes. The 97-Mb total sequence contains 19,099 predicted protein-coding genes—16,260 of which have been interactively reviewed, for an average density of 1 predicted gene per 5 kb (*35*). Each gene has an average of five introns, and 27% of the genome resides in predicted exons. The number of genes is about three times that found in yeast (*2*) and is about one-fifth to one-third the number predicted for humans. As expected from earlier estimates that were based on much smaller amounts of genome sequence, the number of predicted genes is much higher than the number of essential genes that was estimated from classical genetic studies (*10, 36*).

Similarities to known proteins provide a glimpse of the possible function of the predicted genes. Approximately 42% of predicted protein products have distant matches (outside Nematoda); most of these matches contain functional information (*37*). Another 34% of predicted proteins match only other nematode proteins, but only a few

of these have been functionally characterized. The fraction of genes with informative similarities is far lower than the 70% seen for microbial genomes. This may reflect the smaller proportion of nematode genes that are devoted to core cellular functions (*38*), the comparative lack of knowledge of functions involved in building an animal, and the evolutionary divergence of nematodes from other animals studied extensively at the molecular level.

We compared the available protein sets from *C. elegans*, *Escherichia coli*, *Saccharomyces cerevisiae*, and *Homo sapiens* to highlight qualitative differences in the predicted protein sets (*39*) (Fig. 1). Generally, we found that smaller genomes had matches to a larger fraction of their protein sets and larger genomes had higher numbers of matching proteins. As expected from evolutionary relationships, there were substantially more protein similarities found between *C. elegans* and *H. sapiens* than between any other cross-species pairwise comparison. There were also a substantial number of proteins common to *C. elegans* and *E. coli* that were not found in yeast. Similarly, *C. elegans* lacked proteins that were found in both yeast and *E. coli* (*38*).

Genes encoding proteins with distant matches (outside Nematoda) were more likely to have a matching EST (60%) than those without such matches (20%). This observation suggests that conserved genes are more likely to be highly expressed, perhaps reflecting a bias for "housekeeping" genes among the conserved set. Alternatively, genes lacking confirmatory matches may be more likely to be false predictions, although our analyses do not support this (*40*).

We have also used the Pfam protein family database (*41*) to classify common protein domains in the nematode genome. Of the 20 defined domains that occur most frequently (Table 1), the majority are implicated in intercellular communication or in transcriptional regulation. We find comparatively fewer examples of second messenger proteins (for example, 54 G-beta and 3 Src homology 2 domains). This finding supports models in which the same intracellular signaling pathways are used with variant receptors and transcription factors in different cell states.

In addition to the protein-coding genes, the genome contains at least several hundred genes for noncoding RNAs. There are 659 widely dispersed tRNA genes and at least 29 tRNA-derived pseudogenes (*42*). Forty-four percent of the tRNA genes are found on the X chromosome, which contains only 20% of the total sequence. Several other noncoding RNA genes occur in dispersed multigene families. The U1, U2, U4, U5, and U6 spliceosomal RNA genes occur in 14, 21, 5, 12, and 20 dispersed copies, respectively; there are five dispersed copies of signal recognition particle RNA genes, and there are at least four dispersed copies of splice leader 2 (SL2) RNA genes. A striking feature of these dispersed gene families is their high degree of sequence homogeneity. For example, of the 20 U6 RNA genes, 17 are 100% identical to each other. Either gene conversion or recent gene duplications may account for this homogeneity. Several of these RNA genes occur in the introns of protein-coding genes, which may indicate RNA gene transposition. In general, RNA genes in introns do not appear to occur preferentially in the coding orientation of the encompassing transcript, which indicates that these RNA genes are probably expressed independently.

Other noncoding RNA genes occur in long tandem arrays. The ribosomal RNA genes occur solely in such an array at the end of chromosome I. The 5*S* RNA genes occur in a tandem array on chromosome V, with array members separated by SL1 splice leader RNA genes. A few other known RNA genes, such as the small cytoplasmic Ro-associated Y RNA and the *lin-4* regulatory RNA, are found only once in the genome. Some RNA genes that are expected to be present in the genome have yet to be identified, probably because they are poorly conserved at both the sequence and secondary structure level. These include ribonuclease P RNA, telomerase RNA, and 100 or more small nucleolar RNA genes.

Repetitive sequences. Some of the sequence that does not code for protein or RNA is undoubtedly involved in gene regulation or in the maintenance and movement of chromosomes. A significant fraction of the sequence is repetitive, as in other multicellular organisms. We have classified repeat sequences as either local (that is, tandem, inverted, or simple sequence repeats) or dispersed.

Tandem repeats account for 2.7% of the genome and are found, on average, once per 3.6 kb. Inverted repeats account for 3.6% of the genome and are found, on average, once per 4.9 kb. Many repeat families are distributed nonuniformly with respect to genes and, in particular, are more likely to be found within introns than between genes. For example, although only 26% of the genome sequence is predicted to be intronic, it contains 51% of the tandem repeats and 45% of the inverted repeats. The 47% of the genome sequence that is predicted to be intergenic contains only 49% of the tandem repeats and 55% of the inverted repeats. As expected, only a small percentage

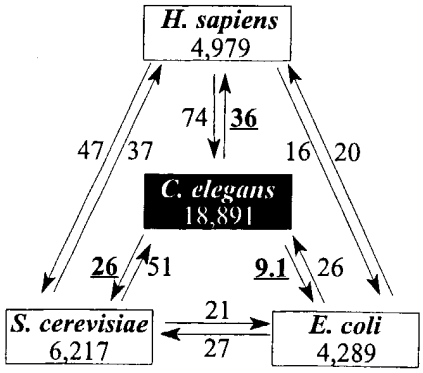


Fig. 1. Percentages of matching proteins resulting from pairwise comparisons (*39*). The organisms and the number of proteins used in the analysis are shown in boxes. For *S. cerevisiae* (a fungus), *C. elegans* (a nematode), and *E. coli* (a bacteria), the numbers reflect proteins that were predicted from an essentially complete genome sequence. The direction of the arrows indicates how the comparison was performed. Numbers that are adjacent to the arrows indicate the percentage of proteins that were found to match. Numbers that are underlined and in bold-faced type indicate the percentage of *C. elegans* proteins that were found to match each of the other organisms.

Table 1. The 20 most common protein domains in *C. elegans* (*41*). RRM, RNA recognition motif; RBD, RNA binding domain; RNP, ribonuclear protein motif; UDP, uridine 5′-diphosphate.

Number	Description
650	7 TM chemoreceptor
410	Eukaryotic protein kinase domain
240	Zinc finger, C4 type (two domains)
170	Collagen
140	7 TM receptor (rhodopsin family)
130	Zinc finger, C2H2 type
120	Lectin C-type domain short and long forms
100	RNA recognition motif (RRM, RBD, or RNP domain)
90	Zinc finger, C3HC4 type (RING finger)
90	Protein-tyrosine phosphatase
90	Ankyrin repeat
90	WD domain, G-beta repeats
80	Homeobox domain
80	Neurotransmitter-gated ion channel
80	Cytochrome P450
80	Helicases conserved C-terminal domain
80	Alcohol/other dehydrogenases, short-chain type
70	UDP-glucoronosyl and UDP-glucosyl transferases
70	EGF-like domain
70	Immunoglobulin superfamily

of the tandem repeats overlaps with the 27% of the genome encoding proteins.

Although local repeat structures are often unique in the genome, others come in families. For example, repeat sequence CeRep26 is the tandemly occurring hexamer repeat TTAGGC, which is seen at multiple sites that are internal to the chromosomes in addition to the telomeres (*22*). CeRep26 and CeRep27 are excluded from introns, whereas other repeat families show a slight positive bias toward introns. The reason for the biased distribution of these repeats is unclear. Furthermore, some repeat families show a chromosome-specific bias in representation. For example, CeRep11, with 711 copies distributed over the autosomes, has only one copy located on the X chromosome.

Altogether, we have recognized 38 dispersed repeat families. Most of these dispersed repeats are associated with transposition in some form (*43*) and include the previously described known transposons of *C. elegans*. However, these repeat elements may not explicitly encode an active transposon (*44*). For example, we have found four new families of the Tc1/mariner type, but these are highly divergent from each other and the other family members; they are probably no longer active in the genome.

In addition to multicopy repeat families, we observe a substantial amount of simple duplication of sequence, that is, segments ranging from hundreds of bases to tens of kilobases that have been copied in the genome. In one case, a segment of 108 kb containing six genes is duplicated tandemly with only 10 sites observed to be different between the two copies. At the left end of chromosome IV, immediately adjacent to the telomere, an inverted repeat is present where each copy of the repeat is 23.5 kb, with only eight different sites found between the two copies. Many cases of shorter duplications are found, which are often separated by tens of kilobases or more that may also contain a coding sequence. These duplications could provide a mechanism for copy divergence and the subsequent formation of new genes. In one example, two 2.5-kb segments, separated by 200 kb, were found to contain genes exhibiting a 98% sequence identity (C38C10.4 and F22B7.5). EST data indicate that both genes are expressed. More commonly, gene duplications are local. In a search for local clusters of duplicated genes, 402 clusters were found distributed throughout the genome (Fig. 2).

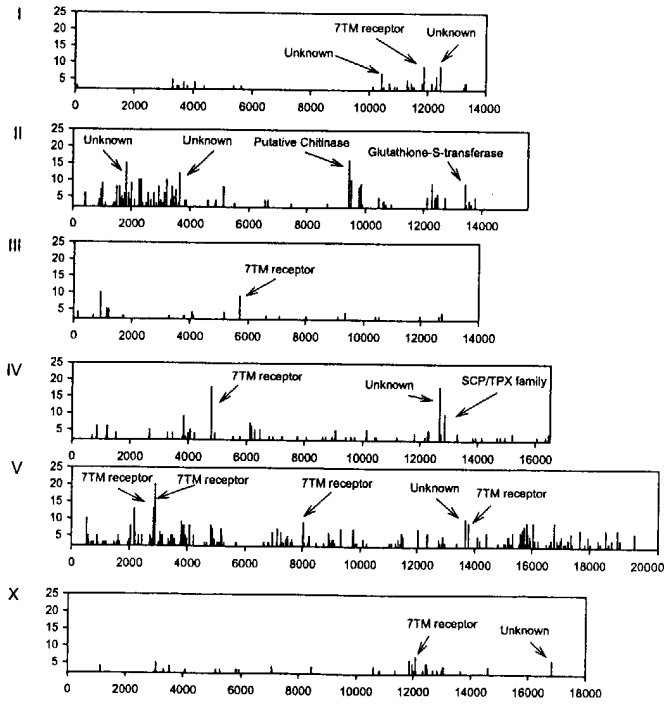


Fig. 2. Locations by chromosome (shown by roman numerals) of local gene clusters. The *x* axis represents the physical distance in kilobases along the chromosomes. The *y* axis represents the size of the clusters. For example, the chitinase cluster on chromosome II contains 17 chitinase-like genes. Local gene clusters were determined by searching for all cases of *N* genes that are similar within a window of 2*N* genes along the chromosomes (for example, three similar genes within a window of six were considered a cluster; clusters were extended until no similar genes could be added). Clusters of $N = 3$ or more were plotted. The criterion for similarity was defined as a BLASTP score of at least 200. ATP, adenosine 5′-triphosphate; TM, transmembrane; Mem. Recep., membrane receptor; SCP/TPX, a family of proteins (SCP, sperm-coating glycoprotein; TPX, Tpx-1, a testis-specific protein).

Chromosome organization. At first sight, the genome looks remarkably uniform; GC content (36%) is essentially unchanged across all the chromosomes, unlike the GC content in vertebrate genomes, such as human, or yeast (*45*). There are no localized centromeres as found in most other metazoa. Instead, the extensive, highly repetitive sequences that are characteristic of centromeres in other organisms may be represented by some of the many tandem repeats found scattered among the genes, particularly on the chromosome arms. Gene density is also fairly constant across the chromosomes, although some differences are apparent, particularly between the centers of the autosomes, the autosome arms, and the X chromosome (Table 2 and Fig. 3).

Striking differences become evident after an examination of other features. Both inverted and tandem repetitive sequences are more frequent on the autosome arms (Fig. 3) than in the central regions of the chromosomes or on the X chromosome. For example, CeRep26 is virtually excluded from the centers of the autosomes (Fig. 3). (The abundance of repeats on the arms is likely to be a contributing factor to the difficulties in cosmid cloning and sequence completion in these regions.) The fraction of genes with

Table 2. Gene density. Autosomes are divided into the genetically defined compartments of the left arm (L), the central cluster region (C), and the right arm (R). The percentage of genes with EST and database matches was determined only from manually inspected genes. Database matches to non-nematode proteins were determined with WUBLASTP ($P \leq 0.001$). Parentheses denote the number of low-scoring predictions thought to be pseudogenes.

Chromosome	Size (Mb)	Protein genes	Density (kb per gene)	tRNA genes	Coding (%)	EST match (%)	Database match (%)
I							
L	3.29	649	5.06	7(2)	21.59	57.0	53.9
C	5.59	1,171	4.77	34(4)	31.65	52.9	52.1
R	4.98	983	5.06	33(2)	25.00	43.4	40.8
II							
L	3.83	1,049	3.65	29(13)	29.00	22.7	26.9
C	7.93	1,719	4.61	38(6)	29.68	49.7	49.8
R	2.96	491	6.03	16(5)	19.89	43.5	39.9
III							
L	3.30	612	5.4	31(14)	20.60	44.2	42.1
C	4.98	1,100	4.52	42(0)	32.21	53.5	53.5
R	4.49	796	5.66	21(3)	23.91	53.1	50.2
IV							
L	5.44	1,050	5.17	38(16)	20.87	39.9	39.7
C	6.51	1,422	4.58	20(3)	29.69	45.7	50.3
R	4.19	622	6.73	26(2)	16.5	36.6	40.7
V							
L	6.19	1,491	4.15	17(4)	27.00	22.0	33.0
C	6.84	1,573	4.34	37(0)	29.40	32.2	43.8
R	7.79	1,018	4.36	152(94)	25.50	19.5	28.8
X	17.22	2,631	6.54	362(33)	19.8	40.9	43.34
Total	95.53	19,141		877(198)			

similarities to organisms other than nematodes tends to be lower on the arms, as does the fraction of genes with EST matches. The difference between autosome arms and central regions is even more obvious in the number of EST matches (*46*). The local gene clusters described above also appear to be more abundant on the arms.

These features, together with the fact that meiotic recombination is much higher on the autosome arms, suggested that the DNA on the arms might be evolving more rapidly than in the central regions of the autosomes. If so, one might expect that the conserved set of eukaryotic genes shared by yeast and *C. elegans* would be largely excluded from the arms. To test this, we identified 1517 proteins in *C. elegans* that are highly similar to yeast genes and plotted their location along the length of the chromosomes (Fig. 3). For four of the five autosomes, the differences in the distribution of core genes are quite striking, with surprisingly sharp boundaries evident. These boundaries appear close to the boundaries in the genetic map that separate regions of high and low rates of recombination (*47*).

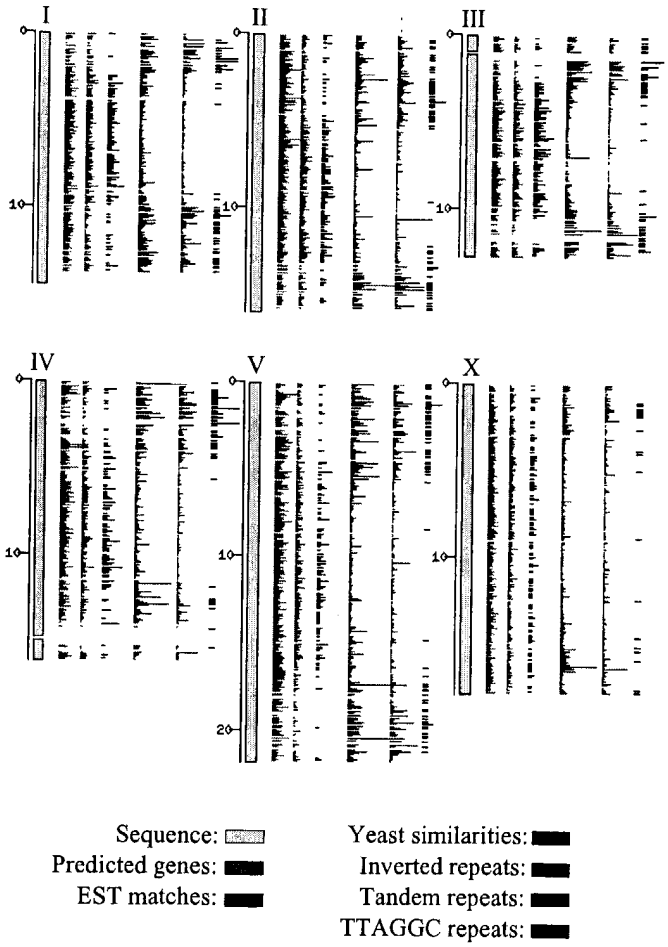


Fig. 3. Distributions of predicted genes; EST matches; yeast protein similarities; and inverted, tandem, and TTAGGC repeats along each chromosome. Gene density varies little along and among the autosomes. On the X chromosome, genes appear at a lower density and are more evenly distributed. In contrast, the frequency of EST matches varies according to their position along the autosomes, indicating a clustering of highly expressed genes. The chromosomal locations of these clusters correlate well with the chromosomal locations of gene products that exhibit significant similarities to yeast proteins (P value of 10^{-9}). For the autosomes, repeat density varies dramatically with chromosomal position and is highest on the arms. The density of inverted and tandem repeats on the X chromosome is more uniform, but similar to the autosomes, TTAGGC repeats tend to be located on the arms. Supplemental information regarding the analysis can be found at www.sciencemag.org/feature/data/c-elegans.shl for a general overview.

Conclusions

There are several reasons for completely sequencing a genome. The first and most simple reason is that it provides a basis for the discovery of all the genes. Despite the power of cDNA analysis and its enormous value in interpreting genome sequence, it is now generally recognized that a direct look at the genome is needed to complete the inventory of genes. Second, the sequence shows the long-range relationships between genes and provides the structural and control elements that must lie among them. Third, it provides a set of tools for future experimentation, where any sequence may be valuable and completeness is the key. Fourth, sequencing provides an index to draw in and organize all genetic information about the organism. Fifth, and most important over time, is that the whole is an archive for the future, containing all the genetic information required to make the organism (the greater part of which is not yet understood). As a resource, the sequence will be used indefinitely not only by *C. elegans* biologists, but also by other researchers for the comparison with and the interpretation of other genomes, including the human genome.

As was already known, the genome of a multicellular organism is very different from that of a microbial organism (and even different from that of a eukaryote such as yeast). It is predominantly noncoding, with genes extended (sometimes over many kilobases) by introns. Rather than acting primarily as the source for a set of protein sequences, the genomic sequence itself remains the primary focus of annotation. There are two reasons for this. First, much information about biological function is located in noncoding sequences; second, current methods of gene identification, both experimental and computational, are not yet accurate and complete enough to provide a definitive set of protein sequences.

If we began again now, would we employ the same approach? Almost certainly (*48*). The clone-based physical map was a critical factor in organizing the project between the two sites. The clones of the map have also been valuable reagents for the research community and continue to be so; the discrete assemblies of cosmids and YACs have been essential to disentangling extensive repeats in many areas. For the numerous small areas that are underrepresented in shotgun assemblies, rare subclones can be readily recovered from the cosmid and YAC subclone libraries.

There are two minor changes that we would make in the sequencing approach. We would add longer insert bacterial clones (for example, bacterial artificial chromosomes) to the map, fingerprinting them in the same manner as cosmids (*48*). Second, we would begin YAC sequencing earlier in the project. That we did not do so on this occasion was for historical reasons [in particular, the availability of the yeast genome sequence (see above)].

How important has the worm project been to the Human Genome Project? Through feedback from many sources, we gather that it has been influential in showing what can be done. Certainly, it is remarkable to look back to 1992, when a paper concerning just three cosmids was published as an important milestone (*10*). Undoubtedly, the worm project has contributed to technology and software development; it is not a unique test-bed, but along with the other genome projects, it has explored ways of increasing scale and efficiency.

Where is the finish line? This publication marks more of a beginning than an end and is another milestone in an ongoing process of the analysis of *C. elegans* biology. It is not very meaningful at any particular point to call genomes of this size finished, because of the inevitable imperfections that will only gradually be resolved. This is true no matter what method of sequencing is adopted. The important thing is not a declaration of completion, but rather the provision of the best possible tools to the users at every stage and a commitment to

maintenance and improvement, through interaction with the user community, as long as that is needed.

References and Notes

1. M. S. Chee *et al.*, in *Cytomegaloviruses*, vol. 154 of *Current Topics in Microbiology and Immunology*, J. K. McDougall, Ed. (Springer-Verlag, Berlin, 1990), pp. 125–169; R. D. Fleischmann *et al.*, *Science* **269**, 496 (1995); C. J. Bult *et al.*, *ibid.* **273**, 1058 (1996). F. R. Blattner *et al.*, *ibid.* **277**, 1453 (1997); S. T. Cole *et al.*, *Nature* **393**, 537 (1998).
2. H. W. Mewes *et al.*, *Nature* **387** (suppl.), 7 (1997); A. Goffeau *et al.*, *Science* **274**, 546 (1996).
3. A. R. Coulson *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 7821 (1986).
4. A. Coulson *et al.*, *Bioessays* **13**, 413 (1991); A. Coulson *et al.*, *Nature* **335**, 184 (1988). The current status of the *C. elegans* physical map is accessible on the World Wide Web (*20*, *21*).
5. The investigations contributing to the *C. elegans* genome project are too numerous to cite. Two early representative publications are I. Greenwald, A. Coulson, J. Sulston, *Nucleic Acids Res.* **15**, 2295 (1987) and S. Ward *et al.*, *J. Mol. Biol.* **199**, 1 (1988).
6. R. Waterston *et al.*, *Nature Genet.* **1**, 114 (1992); W. R. McCombie *et al.*, *ibid.*, p. 124.
7. Y. Kohara, *PNE Protein Nucleic Acid Enzyme* **41**, 715 (1996).
8. R. Okimoto, J. L. Macfarlane, D. O. Clary, D. R. Wohlstenholme, *Genetics* **130**, 471 (1992).
9. D. T. Burke, G. F. Carle, M. V. Olson, *Science* **236**, 806 (1987).
10. J. Sulston *et al.*, *Nature* **356**, 37 (1992).
11. R. Wilson *et al.*, *ibid.* **368**, 32 (1994).
12. M. Vaudin *et al.*, *Nucleic Acids Res.* **23**, 670 (1995).
13. For details of the sequencing process, see (*49*). The process began with the purification of DNA from selected clones of the tiling path. The DNA was sheared mechanically, and after size selection, the resulting fragments were subcloned into M13 or plasmid vectors. Random subclones were selected for sequence generation (the shotgun sequencing approach). Generally, 900 sequence reads per 40 kb of genomic DNA were generated with fluorescent dye–labeled primers or terminators. Bases were determined with PHRED (*50*). An assembly of these random sequences that was generated with PHRAP (*51*) typically resulted in two to eight contigs. Gap closure and resolution of sequence ambiguities were achieved during finishing [using the editing packages GAP (*52*) and CONSED (*53*) and the collection of additional data] through longer reads, directed sequencing reactions using custom oligonucleotide primers on chosen templates, or additional chemistries as required. High-quality finished sequence was analyzed through the use of a suite of programs (including BLAST and GENEFINDER), and the results were stored in ACEDB and submitted to GenBank. Unfinished and finished sequence data were available to investigators by file transfer protocol (ftp) from both sequencing sites (*20*, *21*).
14. C. R. Heiner, K. L. Hunkapiller, S. M. Chen, *Genome Res.* **8**, 557 (1998); L. G. Lee *et al.*, *Nucleic Acids Res.* **20**, 2471 (1992); J. D. Parsons, *Comput. Appl. Biosci.* **11**, 615 (1995).
15. A. A. McMurray, J. E. Sulston, M. A. Quail, *Genome Res.* **8**, 562 (1998).
16. U. J. Kim, H. Shizuya, P. J. de Jong, *Nucleic Acids Res.* **20**, 1083 (1992).
17. S. Cheng, C. Fockler, W. M. Barnes, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 5695 (1994).
18. A clean separation of the YAC DNA from the host chromosomal DNA sometimes required the use of yeast strains in which specific yeast chromosomes are altered in size to provide a window around the YAC that is free of the native chromosomes. L. Hamer, M. Johnston, E. D. Green, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 11706 (1995).
19. S. E. Devine, S. L. Chissoe, Y. Eby, *Genome Res.* **7**, 551 (1997).
20. Available at www.sanger.ac.uk.
21. Available at genome.wustl.edu/gsc/gschmpg.html.
22. C. Wicky *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 8983 (1996).
23. Every region must be sequenced either on each strand or with dye primer and dye terminator chemistry, which extensive comparisons have shown to be at least as reliable as double stranding in revealing and correcting compressions and other base-calling errors. All regions must be represented by reads from two or more independent subclones or from PCR products across the region. If subcloned PCR products are used for a region, three independent clones must be sequenced. Rare exceptions to the general rules of double stranding or alternative chemistry were permitted on the basis of the following. For regions of <50 bases where, despite valid efforts, a finisher is unable to achieve double stranding or double chemistry, the sequence may be submitted (provided the sequence is of high quality and both the finisher and his or her supervisor see no ambiguous bases). When editing, in XGAP, all sequence data must be resolved at the 75% consensus level, either by the collection of additional data or by the editing of poorly called traces. In CONSED, any consensus base with a quality <25% must be manually reviewed to determine if the available data are sufficient to unambiguously support the derived contig sequence. If not, additional data are collected.
24. Each finished sequence is submitted to a series of quality control tests, including verification that all of the finishing rules (*23*) have been followed and a careful verification that the assembly is consistent with all restriction digest information. In addition, every finished sequence undergoes an automatic process of base calling and reassembly with different algorithms than those that were used for the initial assembly and comparison of the resultant consensus by a banded Smith-Waterman analysis [CROSSMATCH (*51*)] against the sequence that was obtained by the finisher. Any discrepancies in assembly or sequence, along with any regions failing to meet finishing criteria, are manually reviewed, and new data are collected as necessary. Only when all discrepancies are accounted for is the sequence passed on for annotation. In turn, if annotation flags any suspicious regions, these are again passed back to the finisher for resolution, either through additional data collection or editing.
25. P. Green and L. Hillier, unpublished software.
26. G. A. Fichant and C. Burks, *J. Mol. Biol.* **220**, 659 (1991); T. M. Lowe and S. Eddy, *Nucleic Acids Res.* **25**, 955 (1997).
27. S. F. Altschul, W. Gish, W. Miller, *J. Mol. Biol.* **215**, 403 (1990); W. Gish, WU-BLAST unpublished software.
28. E. L. L. Sonnhammer and R. Durbin, in *Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology*, R. Altman, D. Brutlag, P. Karp, R. Lathrop, D. Searls, Eds. (AAAI Press, Menlo Park, CA, 1994), pp. 363–368.
29. R. Mott, *Comput. Appl. Biosci.* **13**, 477 (1997).
30. E. L. Sonnhammer, S. R. Eddy, E. Birney, *Nucleic Acids Res.* **26**, 320 (1998); S. R. Eddy *Curr. Opin. Struct. Biol.* **6**, 361 (1996).
31. We identified local tandem and inverted repeats with the programs QUICKTANDEM, TANDEM, and INVERTED (*20*), which search for repeats within 1-kb intervals along the genomic sequence. An index of repeat families used by the project is available at www.sanger.ac.uk/Projects/C_elegans/repeats/.
32. R. Durbin and J. Thierry-Mieg, unpublished software. Documentation, code, and data are available from anonymous ftp servers at lirmm.lirmm.fr/pub/acedb/, ftp.sanger.ac.uk/pub/acedb/, and ncbi.nlm.nih.gov/repository/acedb/.
33. In *C. elegans*, two or more genes can be transcribed from the same promoter, with one gene separated by no more than a few hundred nucleotides from another. In genes undergoing transsplicing, the 5′ exon begins with a splice acceptor sequence, making this 5′ exon more difficult to distinguish from internal exons. This combination of factors may result in two genes being merged into one [T. Blumenthal, *Trends Genet.* **11**, 132 (1995)].
34. We have identified 182 genes possessing alternative splice variants, which are predominately from EST data. Of these, 67 genes produce proteins that differ at their amino termini, 57 genes produce proteins that differ at the carboxyl end, and 59 genes produce proteins that display an internal variation. Of the internal variations, seven genes showed complete exon skipping. Thirty-one genes were found where the 5′ end of an exon had changed, 21 of which resulted in a difference of three or fewer codons. In contrast, of the 24 alternative transcripts that changed the 3′ end of an exon, only 4 resulted in a change of three or fewer codons.
35. Available at www.sanger.ac.uk/Projects/C_elegans/Science98/.
36. R. K. Herman, in *The Nematode* Caenorhabditis elegans, W. B. Wood, Ed. (Cold Spring Harbor Laboratory Press, Plainview, NY, 1988), pp. 17–45; R. Waterston and J. Sulston, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 10836 (1995).
37. These results were obtained with WU-BLAST (version 2.0a13MP), using default parameters and a threshold P value of 10^{-3}. P. Green *et al.*, *Science* **259**, 1711 (1993).
38. S. A. Chervitz *et al.*, *Science* **282**, 2022 (1998).
39. E. L. Sonnhammer and R. Durbin, *Genomics* **46**, 200 (1997).
40. GENEFINDER systematically uses statistical criteria [primarily log likelihood ratios (LLRs)] to attempt to identify likely genes within a region of genomic sequence. Candidate genes are evaluated on the basis of "scores" that reflect their splice site, translation start site, coding potential LLRs, and intron sizes. These scores are normalized by reference to the distribution of combined scores in a simulated sequence as follows: If a given combined score occurs, on average, once in every 10^s nucleotides in simulated DNA, then the corresponding normalized score is set to s. (For example, exons with a normalized score of 5.0 or greater will be found only once in every 100 kb of simulated DNA. With the current reference simulated sequence, which is 1 Mb in length, 6.0 is the maximum normalized score that can occur.) A dynamic programming algorithm is then used to find the set of nonoverlapping candidate genes (on a given strand) that has the highest total score (among all such sets). About 85% of experimentally verified "exon ORFs" (open reading frames containing true exons) in *C. elegans* genes in GenBank have normalized scores above 5.0 (and many of the remaining 15% are initial or terminal exons, which have a single splice site). The fraction of exons with scores >5.0 may be lower for all *C. elegans* genes because of the bias toward highly expressed genes (which often have very high coding segment scores) in the experimentally verified set. However, even for genes in the current verified set that are expressed at moderate to low levels, a majority of exon ORF scores exceed 5.0; this score should be an effective criterion for identifying at least part of most genes. In theory, high-scoring ORFs could arise in other ways. For example, intergenic or intronic regions having abnormal nucleotide composition might appear to have coding segments and occasionally, by chance, may have high-scoring splice sites. So far, there seem to be relatively few such regions in the *C. elegans* genomic sequence. These regions may account for the anomalous orphan exons that we occasionally find. In addition, there are examples where these GENEFINDER-predicted genes fall into clear gene families that are nematode-specific or have only very distant similarity outside the nematodes, for example, chemoreceptor genes (*54*).
41. Pfam is a collection of protein family alignments that were constructed semiautomatically with hidden Markov models within the HMMER package. The collagen and seven transmembrane chemoreceptor data were obtained with unpublished hidden Markov models. The number of seven transmembrane chemoreceptor genes is lower than that found by Robertson (*54*), which could be due to pseudogenes.
42. Putative tRNA pseudogenes are identified by the search program tRNAscan-SE as sequences that are significantly related to a tRNA sequence consensus but do not appear to be likely to adopt a tRNA's canonical secondary structure (*26*). Many higher eukaryotic genomes have mobile, tRNA-derived short interspersed nuclear elements (SINEs). However, because they are few in number, the nematode tRNA

pseudogenes seem more likely to have arisen by some rare event rather than by the extensive mobility that characterizes mobile SINEs [G. R. Daniels and P. L. Deininger, *Nature* **317**, 819 (1985)].
43. A. F. Smit, *Curr. Opin. Genet. Dev* . **6**, 743 (1996).
44. R. F. Ketting, S. E. J. Fischer, R. H. Plasterk, *Nucleic Acids Res.* **25**, 4041 (1997).
45. G. Bernardi, *Annu. Rev. Genet.* **29**, 445 (1995); B. Dujon *et al.*, *Nature* **369**, 371 (1994).
46. The abundance of *C. elegans* ESTs does not directly reflect expression levels, because they are derived from cDNAs in which more abundantly expressed genes were partially selected against (*6*, *7*).
47. T. M. Barnes, Y. Kohara, A. Coulson, *Genetics* **141**, 159 (1995).
48. This approach is also being used for the human genome (Sanger Centre, Washington University Genome Sequencing Center, *Genome Res.*, in press).
49. For methodological details, see (*20*) or (*21*). For biochemical procedures, see R. K. Wilson and E. R. Mardis, in *Genome Analysis: A Laboratory Manual*, B. Birren, E. D. Green, S. Klapholz, R. M. Myers, J. Roskams, Eds. (Cold Spring Harbor Laboratory Press, Plainview, NY, 1997), vol. 1, pp. 397–454. For software packages, see (*20*) or (*21*) and S. Dear *et al.*, *Genome Res.* **8**, 260 (1998); M. Wendl *et al.*, *ibid.*, p. 975; J. D. Parsons, *Comput. Appl. Biosci.* **11**, 615 (1995); and M. Cooper *et al.*, *Genome Res.* **6**, 1110 (1996).
50. B. Ewing, L. Hillier, M. C. Wendl, *Genome Res.* **8**, 175 (1998); B. Ewing and P. Green, *ibid.*, p. 186.
51. P. Green, personal communication.
52. J. K. Bonfield, K. F. Smith, R. Staden, *Nucleic Acids Res.* **23**, 4992 (1995).
53. D. Gordon, C. Abajian, P. Green, *Genome Res.* **8**, 195 (1998).
54. H. M. Robertson, *Genome Res.* **8**, 449 (1998).
55. This work has been supported by grants from the U.S. National Human Genome Research Institute and the UK MRC. We would also like to thank the many members of the *C. elegans* community who have shared data and provided encouragement in the course of this project.

Zinc Fingers in *Caenorhabditis elegans*: Finding Families and Probing Pathways

Neil D. Clarke and Jeremy M. Berg

REVIEW

More than 3 percent of the protein sequences inferred from the *Caenorhabditis elegans* genome contain sequence motifs characteristic of zinc-binding structural domains, and of these more than half are believed to be sequence-specific DNA-binding proteins. The distribution of these zinc-binding domains among the genomes of various organisms offers insights into the role of zinc-binding proteins in evolution. In addition, the complete genome sequence of *C. elegans* provides an opportunity to analyze, and perhaps predict, pathways of transcriptional regulation.

Less than 15 years ago, it was suggested that repeated sequences found in transcription factor IIIA (TFIIIA) of *Xenopus* might fold into structural domains stabilized by the binding of zinc to conserved cysteine and histidine residues (*1*–*3*). Klug and co-workers further noted that "it would not be surprising if the same 30 residue units were found to occur in varying numbers in other related gene control proteins" (*1*). This proposal proved remarkably prescient: *Caenorhabditis elegans*, for example, turns out to have more than 100 such proteins, and the number of domains per protein varies from one to perhaps as many as fourteen. Unanticipated at the time, though, was the fact that the zinc-binding motif found in TFIIIA is just one of many small zinc-binding domains, a number of which are involved in gene regulation. The properties of a few of these domains have been summarized recently (*4*).

Eukaryotes contain a much greater number of proteins with well-characterized zinc-binding motifs than do bacterial and archaeal organisms (Table 1). The complete genome of *Caenorhabditis elegans* (a metazoan), in conjunction with that of *Saccharomyces cerevisiae* (a yeast), presents a special opportunity to examine the range and diversity of these gene families in eukaryotes. Furthermore, because some of these zinc-binding motifs are sequence-specific DNA-binding proteins, the availability of nearly complete sequence information also permits a preliminary analysis of the distribution of potential binding sites within the entire genome. Such analyses may prove to be of value in deducing development control pathways and in more fully defining the characteristics of eukaryotic promoters.

Department of Biophysics and Biophysical Chemistry, The Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA.

The Cys_2His_2 Family

The zinc-stabilized domains of TFIIIA are known as "zinc fingers" or Cys_2His_2 domains. The consensus sequence for this family is (Phe, Tyr)-X-Cys-X_{2-4}-Cys-X_3-Phe-X_5-Leu-X_2-His-X_{3-5}-His (*5*–*7*). In both *C. elegans* and the yeast *S. cerevisiae*, roughly 0.7% of all proteins contain one or more Cys_2His_2 zinc finger domains (Table 1). However, the distribution of these domains within proteins is rather different in the two organisms. In yeast, the majority of zinc finger proteins contain exactly two domains, and only a few (~10%) have more than two. In contrast, there are more zinc finger proteins in *C. elegans* that have three or more Cys_2His_2 domains than there are proteins that have exactly two (Fig. 1) (*8*). On the basis of the sequences of mammalian and *Drosophila* zinc finger proteins, it appears that the distribution of Cys_2His_2 domains among *C. elegans* proteins is typical of multicellular organisms.

The GATA, LIM, and Hormone Receptor Families: Implications for Metazoan Evolution

The GATA domain, the LIM domain, and the DNA-binding domains from nuclear hormone receptors each include a four-cysteine zinc-binding domain that can be clustered into the same structural superfamily, and it is possible that they share a common evolutionary origin (Fig. 2) (*9*, *10*). In addition to the Cys_4 superfamily domain, LIM domains contain a similar LIM-specific $Cys_2HisCys$ zinc motif, whereas the hormone receptors have a second and distinct Cys_4 domain. GATA proteins frequently contain a pair of Cys_4 superfamily domains.

Normalized to the number of genes in their respective genomes, the number of GATA and LIM domain homologs is similar in *C. elegans* and *S. cerevisiae*. In striking contrast, the hormone receptor family is completely absent in yeast but is the largest single family of zinc-binding domains in *C. elegans*. In fact, with over 200 family members, the hormone receptors make up nearly 1.5% of the entire coding sequence of *C. elegans*. The differences in the distribution of nuclear hormone receptors in *C. elegans* and *S. cerevisiae* may be relevant to the evolution of multicellular animals. As has been noted before, the evolution of hormone receptors may have been a key event in the development of cell-cell communication and the origins of multicellularity in the metazoa (*11*).

The ligand-binding domains of the hormone receptors have diverged considerably more than the DNA-binding domains. Applying the same criterion for significance to both the DNA- and ligand-binding domains of the hormone receptor family, only about 10% of the open reading frames (ORFs) that have a DNA-binding domain

Protein families database of alignments and HMMs

SwissPfam entry for Q9XVD7

Description from Swiss-Prot/TrEMBL for Q9XVD7 :

c18d4.9 protein

DUF216 36-324

Key

signal peptide: > pfamA: > Context: > smart: > transmembrane: > low complexity: > coiled coil: > pfamB: Hide

Source	Domain	Start	End
Pfam	**DUF216**	36	324
tmhmm	transmembrane	21	40
tmhmm	transmembrane	64	86
seg	low complexity	69	83
tmhmm	transmembrane	93	115
tmhmm	transmembrane	119	138
tmhmm	transmembrane	151	173
tmhmm	transmembrane	214	236
seg	low complexity	224	236

Overlapping Domains: Change the domain order using the ^ and v buttons. \ the changes by clicking the 'Change order' button.

high priority

pfamA
transmembrane
low complexity

low priority

Change order

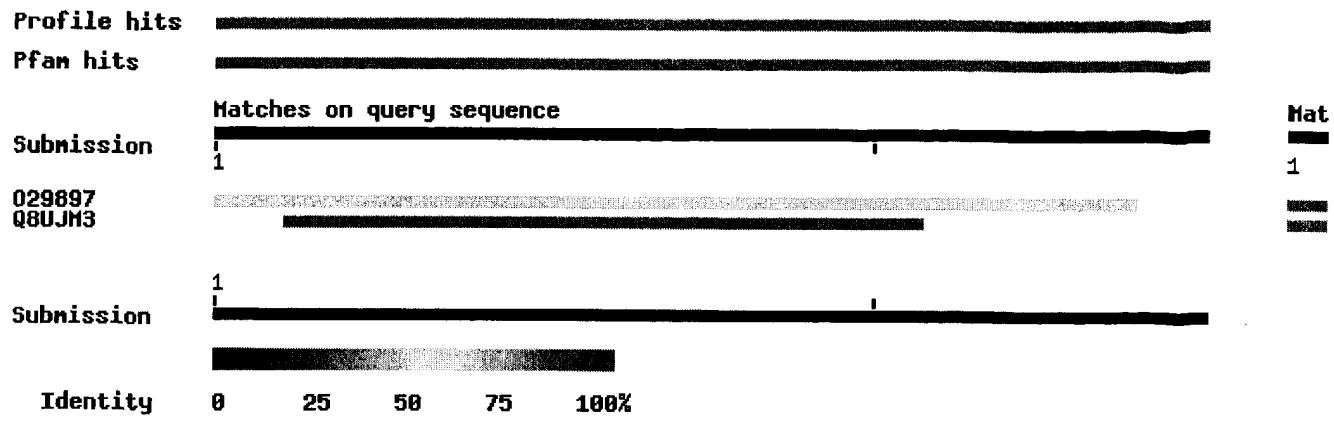
Increase priority

Decrease priority

Hide domain

display domain

Comments or questions on the site? Send a mail to **pfam@sanger.ac.uk**



Alignments

tr O29897 **Conserved hypothetical transmembrane protein [AF0350]** 211 AA
[Archaeoglobus fulgidus]
align

Score = 32.5 bits (69), Expect = 0.72
Identities = 13/22 (59%), Positives = 13/22 (59%), Gaps = 8/22 (36%)

```
Query: 1  WHW--RH-----RIPLQL-AAG 14
          WHW  R       RIPLQL AAG
Sbjct: 25 WHWWLRYNAPCPRIPLQLPAAG 46
```

tr Q8UJM3 **ABC transporter, membrane spanning protein [ATTG]** 802
[Agrobacterium AA
tumefaciens (strain C58 / ATCC 33970)] align

Score = 31.6 bits (67), Expect = 1.3
Identities = 9/10 (90%), Positives = 9/10 (90%), Gaps = 1/10 (10%)

```
Query: 2  HWRHRIPLQL 11
          HWRHR PLQL
Sbjct: 12 HWRHR-PLQL 20
```

Database: EXPASY/UniProt
Posted date: Jun 6, 2004 11:44 AM
Number of letters in database: 476,696,577
Number of sequences in database: 1,498,944

Lambda	K	H
0.344	0.273	2.13

Gapped

Lambda	K	H
0.294	0.110	0.610

Matrix: PAM30

```
Gap Penalties: Existence: 9, Extension: 1
Number of HSP's better than 10.0 without gapping: 0
length of query: 15
length of database: 476,696,577
effective HSP length: 6
effective length of query: 9
effective length of database: 467,702,913
effective search space: 4209326217
effective search space used: 4209326217
T: 16
A: 40
X1: 15 ( 7.4 bits)
X2: 35 (14.8 bits)
X3: 58 (24.6 bits)
S1: 40 (21.7 bits)
S2: 61 (29.1 bits)
```

ExPASy Home page | Site Map | Search ExPASy | Contact us | Proteomics tools | Swiss-Prot

ExPASy Home page | Site Map | Search ExPASy | Contact us | Proteomics tools | Swiss-Prot

Search Swiss-Prot/TrEMBL for q9xvd7

==========

Welcome to the SIB BLAST Network Service

If results of this search are reported or published, please mention that the computation was performed at the SIB using the BLAST network service. The SIB BLAST network service uses a server developed at SIB and the NCBI BLAST 2 software.

In case of problems, please read the online BLAST help.
If your question is not covered, please contact <helpdesk@expasy.org>.

NCBI BLAST program reference [PMID:9254694]:
Altschul S.F., Madden T.L., Schäffer A.A., Zhang J., Zhang Z., Miller W., Lipman D.J. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25:3389-3402(1997).

==========

Query length: 15 AA
Date run: 2004-06-09 22:56:14 UTC+0100 on sib-gm1.unil.ch
Program: NCBI BLASTP 1.5.4-Paracel [2003-06-05]
Database: EXPASY/UniProt
1,498,944 sequences; 476,696,577 total letters

Taxonomic view | NiceBlast view | Printable view

List of potentially matching sequences

Send selected sequences to Clustal W (multiple alignment) Submit Query

Select up to

☐ Include query sequence

	Db AC	Description	Score	E-value
☐	tr O29897	Conserved hypothetical transmembrane protein [AF0350] ...	32	0.72
☐	tr Q8UJM3	ABC transporter, membrane spanning protein [ATTG] [Agr...	32	1.3

Graphical overview of the alignments

Click here to resubmit your query after masking regions matching PROSITE profiles or Pfam HMMs
(? Help) (use ScanProsite for more details about PROSITE matches)

NiceProt View of TrEMBL: Q8UJM3

Note: most headings are clickable, even if they don't appear as links. They link to the user manual or other documents.

Entry information

Entry name	**Q8UJM3**
Primary accession number	**Q8UJM3**
Secondary accession numbers	None
Entered in TrEMBL in	Release 21, June 2002
Sequence was last modified in	Release 21, June 2002
Annotations were last modified in	Release 24, June 2003

Name and origin of the protein

Protein name	**ABC transporter, membrane spanning protein**
Synonyms	None
Gene name	**ATTG** or **ATU5453** or **AGR_PAT_665**
From	Agrobacterium tumefaciens (strain C58 / ATCC 33970) [TaxID: 176299]
Encoded on	Plasmid AT.
Taxonomy	Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales; Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.

References

[1] SEQUENCE FROM NUCLEIC ACID.
MEDLINE=21608550; PubMed=11743193; [NCBI, ExPASy, EBI, Israel, Japan]
Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P., Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Woo L., Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Bovee D. Sr., Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C., Kutyavin T., Levy R., Li M.-J., McClelland E., Palmieri A., Raymond C., Rouse G., Saenphimmachak C., Wu Z., Romero P., Gordon D., Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M., Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M., Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V., Nester E.W.;
"The genome of the natural genetic engineer Agrobacterium tumefaciens C58.";
Science 294:2317-2323(2001).

[2] SEQUENCE FROM NUCLEIC ACID.
MEDLINE=21608551; PubMed=11743194; [NCBI, ExPASy, EBI, Israel, Japan]
Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M., Qurollo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L., Houmiel K., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu

F., Wollam C., Allinger M., Doughty D., Scott C., Lappas C., Markelz B., Flanagan C., Crowell C., Gurson J., Lomo C., Sear C., Strub G., Cielo C., Slater S.;
"Genome sequence of the plant pathogen and biotechnology agent Agrobacterium tumefaciens C58.";
Science 294:2323-2328(2001).

Comments

None

Cross-references

EMBL	AE008966; AAL46140.1; ALT_INIT.[EMBL / GenBank / DDBJ] [CoDingSequence] AE007914; AAK90828.1; -. [EMBL / GenBank / DDBJ] [CoDingSequence]
PIR	AF3215; AF3215.
GO	GO:0046821;Cellular component: extrachromosomal DNA *(inferred from electronic annotation).* GO:0016020;Cellular component: membrane *(inferred from electronic annotation).*
CMR	Q8UJM3; ATU5453.
InterPro	IPR003838; DUF214. Graphical view of domain structure.
Pfam	PF02687; FtsX; 2. Pfam graphical view of domain structure.
ProDom	[Domain structure / List of seq. sharing at least 1 domain]
HOBACGEN	[Family / Alignment / Tree]
ProtoMap	Q8UJM3.
PRESAGE	Q8UJM3.
ModBase	Q8UJM3.
SMR	Q8UJM3; 8D4D192A3A2F7D58.
SWISS-2DPAGE	Get region on 2D PAGE.
UniRef	View cluster of proteins with at least 50% / 90% identity.

Keywords

Plasmid; **Complete proteome**.

Features

None

Sequence information

Length: **802 AA** Molecular weight: **86916 Da** CRC64: **8D4D192A3A2F7D58** [This is a checksum on the sequence]

```
        10         20         30         40         50         60
         |          |          |          |          |          |
MNRTAFQALL SHWRHRPLQL FTLFVGVALA TALWSAVQAI NAEARASYDR AASVLAQNQL

        70         80         90        100        110        120
         |          |          |          |          |          |
DQLVAKDGST ISSNTYARLR RAGLDVSPII EGEHRFGTTR IRLIGIDPLT MPSEGRVLPV

       130        140        150        160        170        180
         |          |          |          |          |          |
GEPSGLIDFM TAPGVMIVSP ATAGTLKDTS GLPIKMAANI PDGAGFVDIS TADRVLHRNG

       190        200        210        220        230        240
```

```
                                                                        |
NLSRIVVSPT QKLDIQAVAT VAPELSLKEA GGRSDVARLT DSFHLNLTAF GFLAFVVGLF

       250        260        270        280        290        300
        |          |          |          |          |          |
IAYSATGLSF EQRRGTFRTL RSLGIPLSSL TTMLVIEITL FALVAGALGV VLGYVVASTL

       310        320        330        340        350        360
        |          |          |          |          |          |
LPGVAATLSG LYGARVAGSL SIRPEWWLTG LGMALVGTWL SSLQHLYRVW RMPILSTAHS

       370        380        390        400        410        420
        |          |          |          |          |          |
RGWTMASVKS LVLQAIAGIF LIALSSLLIW IGSGLLAGFA ILAAFLLGAA FILPPLLALA

       430        440        450        460        470        480
        |          |          |          |          |          |
LRAGERSSRH VLARWFFADT RQQLPGLSLA LMALLLALSA NIGVGTMVSS FRQTFLRWLD

       490        500        510        520        530        540
        |          |          |          |          |          |
QRLAAEVYVT ARDEAEAARL RKWFPDHARA VLPIWSTRGE VSGAQVQIFG VADDPTYRDH

       550        560        570        580        590        600
        |          |          |          |          |          |
WPLILGSATT WDEIASGRGA LVNEQFWRSG NASLGQKIIL PGGWSATVVG IFSDYGNPMG

       610        620        630        640        650        660
        |          |          |          |          |          |
QVIIGINALN QHYPDVEKLR YGLRVAPDDT ADFRRRLIDD FGLPQDNIVD QASLKRQSAA

       670        680        690        700        710        720
        |          |          |          |          |          |
IFEQTFAVTG ALNILTLAVA GFAMFSSLLT LASLRLPQLA PVWALGLRRR DLAWLEFIRA

       730        740        750        760        770        780
        |          |          |          |          |          |
LTLWFVTFVA AIPVGLALAW VLLTIINVEA FGWRLPMILF PWEWVKLGLV ALFAAVISVL

       790        800
        |          |
IPVRQLAKTA PADLLRVFAN ER
```

Q8UJM3 in FASTA format

View entry in original TrEMBL format
View entry in raw text format (no links)
Request for annotation of this TrEMBL entry

BLAST submission on ExPASy/SIB or at NCBI (USA)

Sequence analysis tools: ProtParam, ProtScale, Compute pI/Mw, PeptideMass, PeptideCutter, Dotlet (Java)

ScanProsite, MotifScan

Search the SWISS-MODEL Repository

ExPASy Home page | Site Map | Search ExPASy | Contact us | Swiss-Prot

Hosted by YPRC Korea | Mirror sites: | Australia | Bolivia | Canada | China | Switzerland | Taiwan | USA

The Genome of the Natural Genetic Engineer *Agrobacterium tumefaciens* C58

Derek W. Wood,[1] Joao C. Setubal,[2,4] Rajinder Kaul,[5] Dave E. Monks,[1] Joao P. Kitajima,[2,3] Vagner K. Okura,[2] Yang Zhou,[5] Lishan Chen,[1*] Gwendolyn E. Wood,[1] Nalvo F. Almeida Jr.,[6] Lisa Woo,[1] Yuching Chen,[1†] Ian T. Paulsen,[7] Jonathan A. Eisen,[7] Peter D. Karp,[8] Donald Bovee Sr.,[5] Peter Chapman,[5] James Clendenning,[5] Glenda Deatherage,[5] Will Gillet,[5] Charles Grant,[5] Tatyana Kutyavin,[5] Ruth Levy,[5] Meng-Jin Li,[5] Erin McClelland,[5] Anthony Palmieri,[5] Christopher Raymond,[5] Gregory Rouse,[5] Channakhone Saenphimmachak,[5] Zaining Wu,[5] Pedro Romero,[8] David Gordon,[9] Shiping Zhang,[10] Heayun Yoo,[10] Yumin Tao,[11] Phyllis Biddle,[10] Mark Jung,[10] William Krespan,[10] Michael Perry,[10] Bill Gordon-Kamm,[11] Li Liao,[10] Sun Kim,[10] Carol Hendrick,[11] Zuo-Yu Zhao,[11] Maureen Dolan,[10] Forrest Chumley,[10‡] Scott V. Tingey,[10] Jean-Francois Tomb,[10] Milton P. Gordon,[12] Maynard V. Olson,[5] Eugene W. Nester[1,13§]

The 5.67-megabase genome of the plant pathogen *Agrobacterium tumefaciens* C58 consists of a circular chromosome, a linear chromosome, and two plasmids. Extensive orthology and nucleotide colinearity between the genomes of *A. tumefaciens* and the plant symbiont *Sinorhizobium meliloti* suggest a recent evolutionary divergence. Their similarities include metabolic, transport, and regulatory systems that promote survival in the highly competitive rhizosphere; differences are apparent in their genome structure and virulence gene complement. Availability of the *A. tumefaciens* sequence will facilitate investigations into the molecular basis of pathogenesis and the evolutionary divergence of pathogenic and symbiotic lifestyles.

Agrobacterium tumefaciens is an α-proteobacterium of the family Rhizobiaceae and a member of the diverse *Agrobacterium* genus. A ubiquitous soil organism and etiological agent of the plant disease crown gall (*1*), *A. tumefaciens* infects more than 90 families of dicotyledonous plants, resulting in major agronomic losses (*2, 3*). The gall results from the transfer, integration, and expression of a discrete set of genes (T-DNA) located on the tumor-inducing (Ti) plasmid. Expression of these genes leads to biosynthesis of plant growth hormones as well as a bacterial nutrient source called opines (*4*). The processing and transfer of the T-DNA is mediated by the Ti plasmid virulence (*vir*) genes, and several virulence determinants initially characterized in *A. tumefaciens* have been found in plant symbionts and animal pathogens (*5–7*).

The genes within the T-DNA can be replaced by any DNA sequence, making *A. tumefaciens* an ideal vehicle for gene transfer and an essential tool for plant research and transgenic crop production. The research and commercial potential of *A. tumefaciens* has been broadened under laboratory conditions to include the transfer of T-DNA to recalcitrant plants, fungi (*8*), and human cells (*9*).

A. tumefaciens shares a similar habitat and close evolutionary relationship with the nitrogen-fixing symbionts of the Rhizobiaceae (*10*). Indeed, the introduction of a symbiotic plasmid from *Rhizobium phaseoli* into *A. tumefaciens* results in the weak but measurable formation of nitrogen-fixing root nodules (*11*), suggesting a shared genetic background. The recent publication of the genome sequences of two Rhizobiaceae, *Sinorhizobium meliloti* (*12*) and *Mesorhizobium loti* (*13*), allowed a genome-wide comparison with *A. tumefaciens*. We present the results of this comparison as well as a detailed analysis of the genome of *A. tumefaciens* strain C58 (*14, 15*).

General features of the genome. The 5.67-Mb genome of *A. tumefaciens* C58 (*16*) comprises four replicons (*17*): a circular chromosome, a linear chromosome, and the AtC58 and TiC58 plasmids (Table 1 and Fig. 1). The genome contains 5419 predicted protein-coding genes (*14*), of which we have assigned a putative function to 3475 (64.1%). The remaining 1944 genes (35.9%) include 1236 conserved hypothetical genes (22.8%) whose predicted products are similar to proteins of unknown function in other genomes, and 708 hypothetical genes (13.1%) with no significant matches in the sequence databases (Table 1). Our analysis assigns the *A. tumefaciens* genes to 501 paralogous families containing from 2 to 206 members (*14*). The two largest families are composed of genes belonging to the adenosine triphosphatase (ATPase) and membrane-spanning components of the ATP binding cassette (ABC) transport family.

The overall GC content of the *A. tumefaciens* genome is 58%. The TiC58 plasmid has two regions of distinctive GC content: the T-DNA (46%) and the *vir* region (54%) (Fig. 1). Low GC content was noted previously in the T-DNA of a related Ti plasmid (*18*). Reduced GC content (53%) is also seen within a 24-kb segment of pAtC58 (AT island, Fig. 1). This region includes 17 conserved hypothetical or hypothetical genes, an ATP-dependent DNA helicase, and an insertion sequence (IS) element. These genes are flanked by a phage integrase and a second IS element. The genes in these three regions have a distinct codon usage as compared to the rest of the genome, consistent with their recent evolutionary acquisition (*14*).

The genome contains 53 transfer RNAs (tRNAs) that represent all 20 amino acids (Table 1). These tRNAs are distributed unevenly between the circular and linear chromosomes. Transfer RNA species corresponding to the most frequently represented ala-

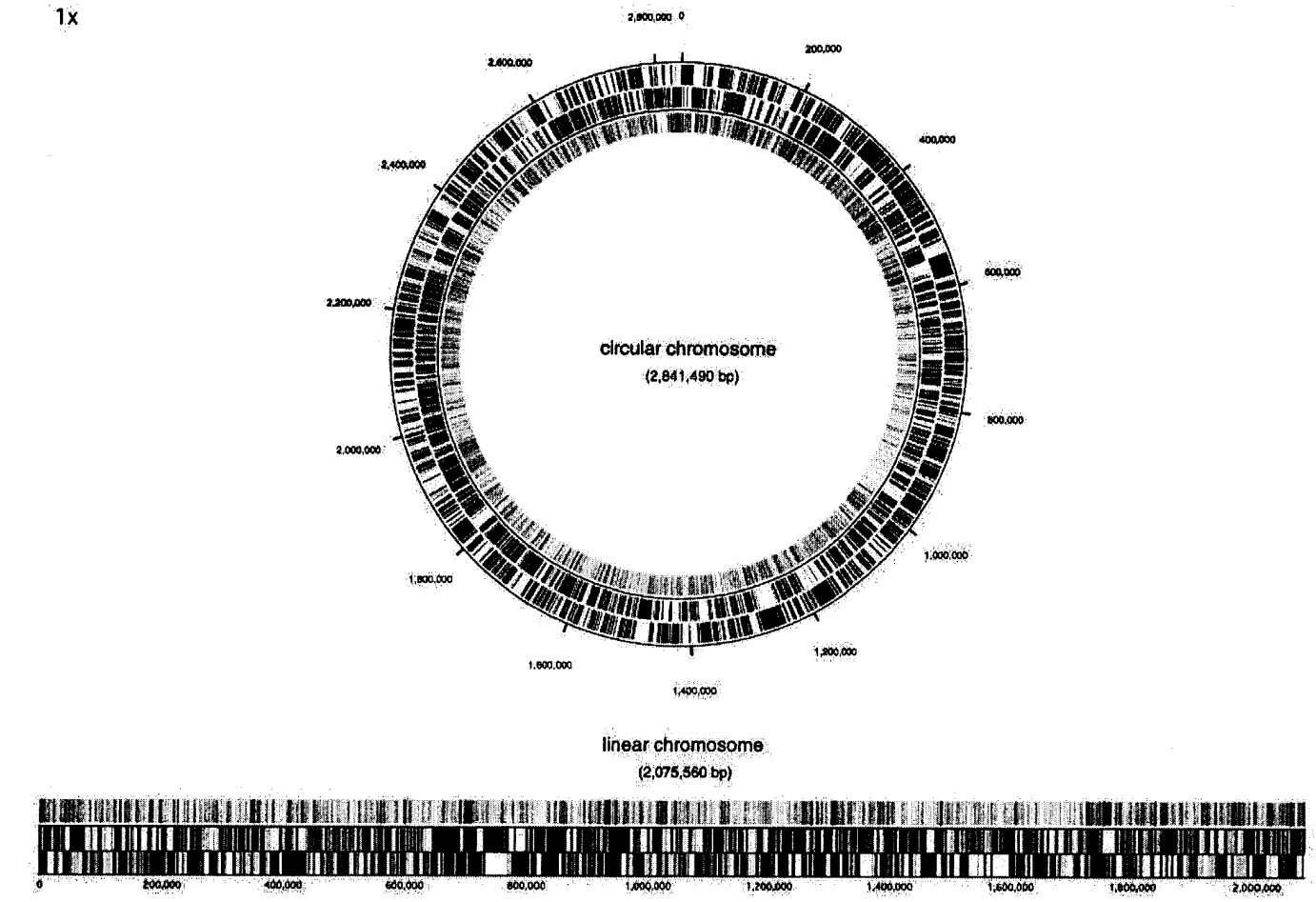
[1]Department of Microbiology, University of Washington, 1959 NE Pacific Street, Box 357242, Seattle, WA 98195, USA. [2]Bioinformatics Laboratory, Institute of Computing, [3]Center for Molecular Biology and Genetic Engineering (CBMEG), University of Campinas, CP 6176, Campinas SP 13083-970, Brazil. [4]Department of Genome Sciences, University of Washington, Box 357730, Seattle, WA 98195, USA. [5]Genome Center, University of Washington, Fluke Hall on Mason Road, Box 352145, Seattle, WA 98195, USA. [6]Department of Computing and Statistics, Federal University of Mato Grosso do Sul, CP 549, Campo Grande MS 79070-900, Brazil. [7]The Institute for Genomic Research, 9712 Medical Center Drive, Rockville, MD 20850, USA. [8]Bioinformatics Research Group, SRI International, 333 Ravenswood Avenue, Menlo Park, CA 94025, USA. [9]Howard Hughes Medical Institute, University of Washington, Box 357730, Seattle, WA 98195, USA. [10]E. I. du Pont de Nemours Company, 1 Innovation Way, Newark, DE 19714, USA. [11]Pioneer Hi-Bred International Inc., 7300 NW 62nd Avenue, Post Office Box 1004, Johnston, IA 50131, USA. [12]Department of Biochemistry, University of Washington, 1959 NE Pacific Street, Seattle, WA 98195, USA. [13]Department of Botany, University of Washington, 1959 NE Pacific Street, Box 355325, Seattle, WA 98195, USA.

*Present address: Department of Pathology, University of Washington, Box 357470, Seattle, WA 98195, USA.

†Present address: Gene Function & Target Validation, Celltech R&D Inc., Bothell, WA 98021, USA.

‡Present address: Department of Plant Pathology, Kansas State University, 113 Waters Hall, Manhattan, KS 66506, USA.

§To whom correspondence should be addressed. E-mail: gnester@u.washington.edu


1x
circular chromosome
(2,841,490 bp)
linear chromosome
(2,075,560 bp)
10x
pTiC58
(214,233 bp)
T-DNA
vir region
5x
pAtC58
(542,779 bp)
AT island

nine, glutamine, and valine codons are found only on the linear replicon. The genome contains 25 predicted IS elements representing eight different families (*14*). The largest is the IS3 family comprising 10 IS elements. The IS elements are not equally distributed among the replicons but are located preferentially on the linear chromosome and pAtC58 (Table 1). The adjacent *virH1* and *virH2* genes of the Ti plasmid, encoding p450 mono-oxygenases (*19*), are flanked by IS elements, which suggests that they arrived in *A. tumefaciens* as part of a compound transposon. Twelve genes of probable phage origin were identified, most of which are on the circular chromosome (Table 1). Many of these genes cluster in two discrete regions and thus may represent prophage remnants. None of these clustered phage-related genes are shared with *S. meliloti*, which implies that they were lost from *S. meliloti* or entered the *A. tumefaciens* genome after these organisms evolutionarily diverged.

Phylogeny and whole-genome comparison. A comparison with all sequenced organisms reveals that the *A. tumefaciens* proteome is most similar to that of two rhizobial species, *S. meliloti* and *M. loti* (*14*). This result was obtained by cataloging top BLAST hits of predicted *A. tumefaciens* proteins and by classifying predicted proteins into clusters of orthologous groups (Fig. 2A) (*20*). Of the two rhizobial species, the *A. tumefaciens* proteome is most similar to that of *S. meliloti*. Phylogenetic analyses of broadly conserved proteins indicate that this similarity results from *A. tumefaciens* and *S. meliloti* sharing a recent common ancestor, and not from gene loss or branch rate variation (Fig. 2, B and C).

Sinorhizobium meliloti has a circular chromosome (3.65 Mb) and two plasmids (1.68 Mb and 1.35 Mb), with a total genome size 1.1 Mb larger than that of *A. tumefaciens* (*12*). The circular chromosomes of these organisms show extensive nucleotide colinearity and gene order conservation (Fig. 3) (*14*). Previously, such extensive colinearity has only been seen between members of the same genus. Chromosome-wide conservation of gene order is less pronounced between *S. meliloti* and *M. loti* (*14*). The comparison of the circular chromosomes of *A. tumefaciens* and *S. meliloti* also reveals major rearrangements near the putative replication origin and termini (Fig. 3, regions A and B). Similarly located rearrangements are commonly seen between closely related bacteria (*21*).

A comparison of the other replicons of *A. tumefaciens* with all replicons of *S. meliloti* reveals a mosaic pattern of ortholog distribution (Table 2 and Fig. 1). These orthologs are distributed across the *A. tumefaciens* elements as individual genes and small regions of gene order conservation. Two regions of the linear replicon exhibit extensive conservation of gene order with a segment of the *S. meliloti* chromosome (Fig. 3, region C). The first comprises 46 genes (44 kb) and the second contains 65 genes (89 kb). These regions are partially conserved in the *M. loti* chromosome. The large number of orthologs and the lack of extensive gene order conservation suggest that the smaller *A. tumefaciens* replicons underwent substantial rearrangement since the organisms diverged. This finding is consistent with differential evolutionary pressures acting on these elements. The nonorthologous genes, many of which are seen on the Ti plasmid, reflect lineage-specific gene loss or acquisition from other species. Taken together, these data support the recent evolutionary divergence of *A. tumefaciens* and *S. meliloti*.

Genus-specific genes. Comparison of the genomes of *A. tumefaciens*, *S. meliloti*, and *M. loti* identified genes in each organism that likely contribute to genus-specific biology (*14*). Of the 5419 predicted *A. tumefaciens* proteins, 853 (16%) are not found in these other organisms. Of these, 97 have an assigned function, whereas 756 are hypothetical or conserved hypothetical. The predicted products of these genes are diverse and include proteins involved in cellulose production, plasmid maintenance, cell growth, transcriptional regulation, and cell wall synthesis. Several additional proteins are predicted to catabolize plant cell wall materials, sugars, and exudates. These include polygalacturonases, a glycosidase, an endoglucanase, a myo-inositol catabolism protein, and a cell wall lysis–associated protein. Additional genes, predictably found on the Ti plasmid, include those encoding virulence, T-DNA, and conjugal transfer–associated proteins. With 756 open reading frames (ORFs) yet to characterize, much remains to be elucidated regarding the genetic distinction between *A. tumefaciens* and its Rhizobiaceae relatives.

Linear chromosome. Linear replicons, the predominant genetic element in eukaryotes, have been identified in only a few prokaryotes. These include members of the genera *Borrelia* and *Streptomyces* (*22, 23*). Although sequence analysis did not reveal distinct features associated with terminal secondary structures, Goodner *et al.* found that the termini of this replicon are covalently linked (*15*). This covalent linkage did not prevent nearly complete sequencing of the replicon termini as confirmed by Southern analysis (*14*). Proteins associated with the maintenance of linear ends in other systems, such as telomerases or the *Streptomyces tpg* proteins (*24*), are absent in *A. tumefaciens*. One notable feature of the replicon termini is the presence of IS elements near each end. The evolutionary origin of this replicon awaits investigation, as does the mechanism that *A. tumefaciens* uses to maintain it in a linear form.

There are 1882 protein-coding genes on the linear replicon, including those encoding ribosomal and DNA replication proteins, as

Fig. 1. (facing page) Schematic representation of the *A. tumefaciens* genome. Chromosomes are drawn to scale with plasmids represented at 5× or 10× magnification, as indicated. The outer two bands indicate opposing transcriptional orientations of predicted genes. Colors indicate orthology to proteins in the *S. meliloti* replicons: blue, chromosome; green, pSymA; gold, pSymB; red, nonorthologous. The inner circle depicts GC content for each coding region, with lower GC content indicated by darker shading. The *vir* and T-DNA regions of pTiC58 and the AT island of pAtC58 are indicated. Orthologs were identified by comparison of predicted proteins for each *A. tumefaciens* replicon with the genome of *S. meliloti*. Two proteins were considered orthologous if their BLASTP alignment covered at least 60% of each protein at an expect value of less than or equal to 10^{-5}. Proteins that did not match these criteria were considered nonorthologous (*14*).

Table 1. General features of the *A. tumefaciens* C58 genome.

Feature	Circular	Linear	pAtC58	pTiC58	Total
Size (bp)	2,841,490	2,075,560	542,779	214,233	5,674,062
G+C content (%)	59.4	59.3	57.3	56.7	58.13
Protein-coding genes					
Assigned function	1715	1286	333	141	3475 (64.1%)
Conserved hypothetical	710	353	128	45	1236 (22.8%)
Hypothetical	364	243	89	12	708 (13.1%)
Total	2789	1882	550	198	5419
Average ORF size (bp)	892	988	843	925	922
Coding (%)	87.9	89.9	85.4	85.5	88.3
Regulators (%)	7.7	10.4	11.8	5.1	9.0
ABC transport	47	80	20	6	153
RNA					
rRNA	2	2	0	0	4
tRNA	40	13	0	0	53
tmRNA	1	0	0	0	1
IS elements	2	10	10	2	24
Phage-related	10	1	1	0	12

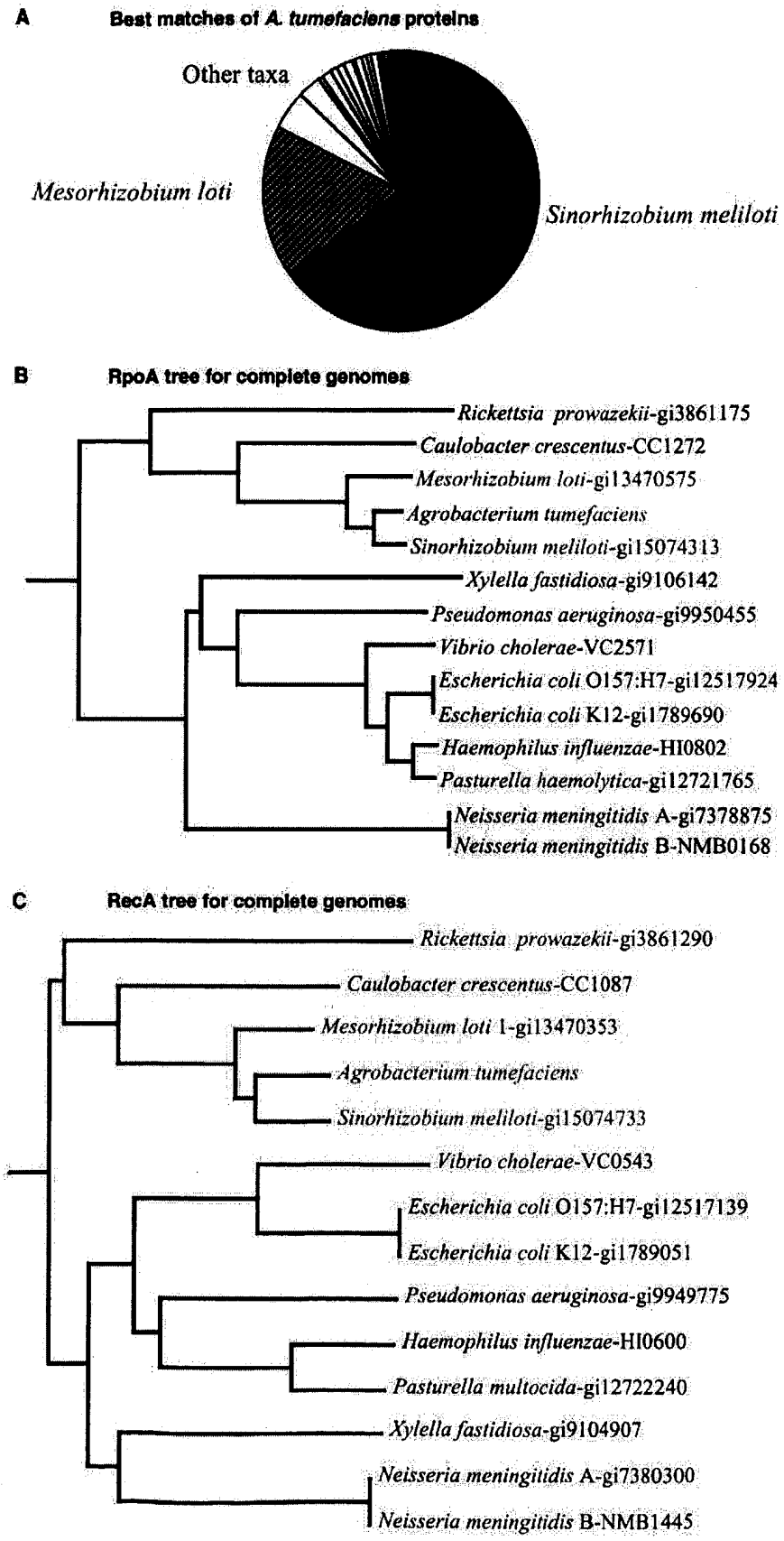


Fig. 2. Comparisons with fully sequenced genomes. (**A**) Distribution of best hits based on a comparison of predicted proteins of *A. tumefaciens* with proteins from all published genomes. (**B** and **C**) Phylogenetic trees generated using two broadly conserved proteins. The trees were generated using PAUP distance methods and a distance calculation based on PAM matrices (*14*).

well as 21 complete metabolic pathways. The presence of these genes confirms the chromosomal identity of this replicon. Additional features, however, resemble those traditionally associated with plasmids. For example, genes whose products are similar to the conjugative proteins TraA, MobC, and TraG are present, although an *oriT* is not apparent. Further, an intact and highly conserved *repABC* operon, the definitive element of the RepABC-type replicator family of circular plasmids, is located near the center of the linear chromosome. The presence of this operon, coupled with a colocalized GC-skew inversion, indicates a bidirectional plasmid-like mode of replication. If experimentally verified, this replication mechanism would prove unique among known linear replicons.

Plasmid replication and transfer. Replication of both pTiC58 and pAtC58 is mediated by RepABC-type systems commonly found in plasmids of the Rhizobiaceae. It is likely that the origin of replication for these plasmids is adjacent to the *repC* gene (*25*). In contrast to the pSymB plasmid of *S. meliloti* (*12*), both *A. tumefaciens* plasmids contain all necessary machinery for conjugation and do not contain essential genes. A new conjugal transfer system belonging to the Type IV secretion family (AvhB) (*26*) was identified on pAtC58. In contrast to the tight control of Ti plasmid conjugal transfer mediated by specific opines that activate quorum sensing (*27*), the conjugal transfer of pAtC58 appears to be constitutive.

Transport. Transporters constitute 15% of the *A. tumefaciens* genome, 87% of which are found on the chromosomes (*14*). These systems are predicted to confer broad capabilities for the transport of common nutrients found in the rhizosphere, including sugars, amino acids, and peptides. In addition, there are 11 LysE/RhtB amino acid efflux proteins, almost double the number seen in any bacterium outside of the Rhizobiaceae (*12, 28*). These transporters may function in the export of homoserine lactones or other signal molecules. There are also a large number of high-affinity tripartite ATP-independent periplasmic (TRAP) dicarboxylate transporters (*29*). Our analyses indicate that *A. tumefaciens* and the other sequenced members of the Rhizobiaceae have similar transport capabilities.

Like both *S. meliloti* and *M. loti*, *A. tumefaciens* has an abundance of ABC transporters, constituting 60% of its total transporter complement. There are 153 complete systems plus additional "orphan" subunits. The number of ABC transporters found in these organisms is greater than that found in any sequenced eukaryote and more than double the number found in any sequenced bacterium (*28, 30*). Predicted substrates of these ABC transport systems include sugars (53 systems), amino acids (29 systems), and pep-

tides (25 systems). Other organisms with large ABC transporter complements include photosynthetic bacteria such as *Synechocystis* PCC6803 and organisms that lack a tricarboxylic acid (TCA) cycle and an electron transfer chain, like the mycoplasmas and *Thermotoga maritima* (*28*). The generation of large ATP pools in these organisms, via photosynthesis or F-type ATPases, may explain their preference for ATP-driven transport. In contrast, the preference for ABC transporters in *A. tumefaciens* may reflect a need for high-affinity uptake systems for the acquisition of nutrients in the highly competitive soil and rhizosphere environments.

Regulation. Bacteria that inhabit diverse environments tend to have large complements of regulatory genes (*31*). Consistent with this, regulatory genes constitute a substantial proportion (9%) of the *A. tumefaciens* genome (Table 1) (*14*). This regulatory capacity likely facilitates survival of *A. tumefaciens* within the dynamic soil and rhizosphere environments. The genome encodes 11 extracellular sigma factors, proteins implicated in stress responses in other organisms (*32*). In addition, although several LuxR family motifs are evident, only one previously identified acyl-homoserine lactone synthase (*traI*) known to be involved in quorum sensing was detected (*4*). Several proteins are similar to eukaryotic signal transduction proteins rarely found in bacteria, including four regucalcin-like calcium-binding regulators and a serine-threonine kinase. As is true of other α-proteobacteria, no *rpoS* gene was identified. However, *A. tumefaciens* does have a homolog of the HF-1 protein known to regulate stationary phase and oxidative stress responses in *Escherichia coli* and *Brucella abortus* (*33*).

Our analysis identified numerous nucleotide cyclases in the plant symbionts *S. meliloti* and *M. loti* (25 and 12, respectively) and in the evolutionarily distinct human pathogen *Mycobacterium tuberculosis* (12). These cyclases are rarely found in other bacterial genomes. The nucleotide cyclases in *S. meliloti* have been noted previously and were postulated to function in signal transduction (*12*). Contrary to our expectation, there are only three nucleotide cyclases in *A. tumefaciens*. It is unclear why the nitrogen-fixing plant symbionts share similarly large numbers of nucleotide cyclases with a human pathogen, whereas few such genes are found in the evolutionarily related *A. tumefaciens*.

Attachment, cell surface, and secretion. The initial interaction of *Agrobacterium* with its plant hosts is mediated by several attachment-related genes (*34*). These include the *chvA*, *chvB*, *exoC*, and cellulose synthesis genes as well as the pAtC58-localized *att* region. Several additional genes encode proteins similar to adhesins in mammalian pathogens, including BfrA of *E. coli* and PsaA of *Streptococcus pneumoniae*.

Pili are extracellular appendages often required for bacterial association with their hosts. Although only the pilus encoded by the *virB* operon has been experimentally confirmed (*35*), the *trb* operon required for Ti plasmid conjugation likely produces a pilus (*36*). The *avhB* and *ctp* clusters, identified by our analyses, may also produce pili. Additional surface components include exopolysaccharides, lipopolysaccharides, and capsular polysaccharides, whose biosynthetic genes are primarily located on the linear chromosome. Such surface polysaccharides are commonly involved in invasion, growth, and survival of plant-associated bacteria.

Five protein secretion systems are found among Gram-negative bacteria (*37*), at least three of which are represented in *A. tumefaciens*. These include four potential type I secretion systems. Although components of the main terminal branch of type II secretion appear to be absent, the Sec system for protein secretion across the inner membrane is intact. There is, however, a type IV pilus biogenesis system with components similar to those of type II secretion systems. Similar to *S. meliloti* (*12*), no type III secretion system was identified. *Agrobacterium tumefaciens* encodes three type IV secretion systems: VirB, Trb (*27*), and AvhB. The genome also contains the twin arginine targeting system, Tat/Mtt (*38*).

Virulence. To date, most virulence determinants of *A. tumefaciens* have been found on the Ti plasmid. Other than the *virB* operon, these genes are not found in *S. meliloti*. The TiC58 plasmid contains a single T-DNA region, in contrast to the two found in a number of other strains (*18*), and the 25–base pair (bp) border regions that delineate the T-DNA are not present elsewhere in the genome.

The availability of the genome sequence has enabled the identification of genes whose products are similar to plant pathogen virulence proteins required for host cell wall degradation. These include pectinase (*kdgF*), ligninase (*ligE*), and xylanase as well as regulators of pectinase and cellulase production (*pecS/M*); *A. tumefaciens* may use such enzymes to breach the cell wall of its host before T-DNA transfer.

In addition, we have identified numerous orthologs of animal virulence genes. Examples include those involved in host survival, such as the *bacA* locus of *Brucella* (*39*) and two members of the widely conserved HtrA family of serine proteases implicated in response to oxidative stress in *Salmonella* and *Yersinia* (*40*). Interestingly, a *bacA* homolog is involved in *S. meliloti* symbiosis (*41*). Invasion-related homologs include the *ialA* and *ialB* genes of *Bartonella henselae* (*42*) as well as five hemolysin-like proteins with associated type I secretion systems. The highly conserved *mviN* gene, implicated in *Salmonella* virulence (*43*), is also present.

Metabolism. *Agrobacterium tumefaciens* grows on minimal medium and therefore possesses all pathways required for pro-

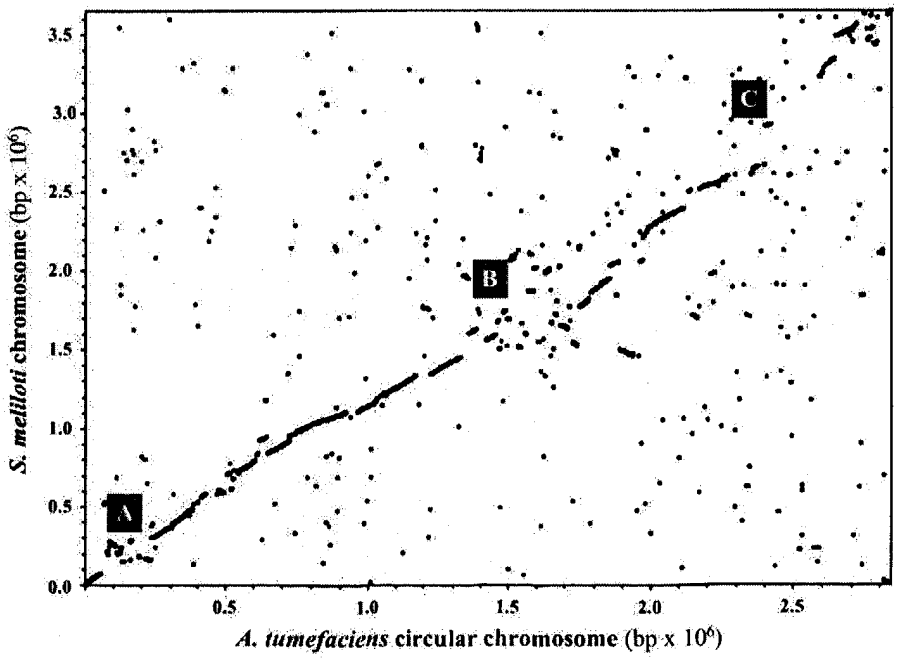


Fig. 3. Alignment of the proteomes of the *S. meliloti* chromosome and the *A. tumefaciens* circular chromosome. Each point in the figure is a bidirectional best hit. These hits were obtained by pairwise BLASTP searches of predicted *A. tumefaciens* proteins against those of *S. meliloti* with a maximum expect value of 10^{-4} (*14*). Putative origins (region A) and termini (region B) of replication are indicated, as well as a sizable region lacking colinearity (region C).

totrophic growth, an observation confirmed by our computational pathway analysis (*14*). These metabolic pathways are dispersed among the four replicons. Unlike their organization in *E. coli*, most genes of these pathways are not tightly clustered, which suggests that they are not present in operons. We identified pathways for the synthesis of all 20 amino acids as well as numerous enzyme cofactors. At least one nonribosomal protein synthesis system for the production of polyketides was identified. Encoded energy metabolism pathways include glycolysis, TCA cycle, and Entner-Doudoroff. *Agrobacterium tumefaciens* can catabolize 17 amino acids, including *S*-adenosylhomocysteine and 4-hydroxyproline. Pathways for use or degradation of plant metabolites typically found in the rhizosphere were also detected. These include sugars such as glucose, fructose, sucrose, ribose, xylose, xylulose, and lactose as well as compounds such as myo-inositol, hydantoin, urea, and glycerol. The capacity to metabolize glucuronate, galactonate, galactarate, gluconate, ribitol, glycogen, quinate, L-idonate, creatinine, stachydrine, ribosylnicotinamide, and 4-hydroxymandelate was also detected. Chemotaxis systems responding to many of these compounds are present in *A. tumefaciens* (*44*).

Agrobacterium tumefaciens encodes a variety of proteins that may protect against toxic compounds in the environment. Examples include four cytochrome p450s, two of which have been previously identified. One of these has been shown to modify ferrulic acid, an inducer of the *vir* genes (*45*). These highly oxidative enzymes may also detoxify or modify plant-derived compounds, including phytoalexins (*46*) and protocatechuate, and xenobiotics such as 1,2-dichloroethane, cyanate, 1,4-dichlorobenzene, and octane. In addition, antibiotic resistance genes targeted against tetracycline (*47*) and chloramphenicol are present.

Many components of nitrogen metabolism are conserved between *A. tumefaciens* and the nitrogen-fixing symbionts *S. meliloti* and *M. loti*. Examples include components of the nitrogen regulation (Ntr) system such as *ntrBC*, *ntrXY*, *ntrA*, *glnE*, *glnD*, *glnB*, and *glnK*. *Agrobacterium tumefaciens* harbors seven glutamine synthetase (GS) genes, which encode GS types I, II, and III. The presence of multiple GS genes may relate to the observation that *A. tumefaciens* requires high concentrations of glutamate for optimal growth. Other members of the Rhizobiaceae also contain multiple GS genes. In addition, *A. tumefaciens* has a gene predicted to encode the large hexameric adenosine monophosphate (AMP)–dependent glutamate dehydrogenase (*48*), but a gene encoding the AMP-independent glutamate dehydrogenase was not identified. In contrast, *S. meliloti* and *M. loti* contain both genes. The *A. tumefaciens* linear chromosome carries denitrification genes, including a periplasmic dissimilatory nitrate reductase (*nap*), nitrite reductase (*nir*), and nitric oxide reductase (*nor*), but lacks nitrous oxide reductase (*nos*). In contrast, all of these genes are present in *S. meliloti*. Nitrate transport genes are also located on the linear chromosome. Although *A. tumefaciens* is considered an aerobe, the existence of these genes implies that it could use nitrate as an electron acceptor under anaerobic conditions. As expected, *A. tumefaciens* lacks the subunits of nitrogenase and its cofactors. Most of the *nod* genes are also absent, except for three genes similar to those involved in nod factor production, *nodL*, *nodX*, and *nodN*.

Conclusions. The combination of a linear and a circular chromosome is found in only a few members of the genus *Agrobacterium* (*49*). This observation represents a key evolutionary distinction between *A. tumefaciens* and *S. meliloti*. On the basis of 16*S* ribosomal DNA phylogenetic analyses, it has been proposed that the genus *Agrobacterium* be reclassified into the genus *Rhizobium* (*10*). Combining what has been elucidated regarding genome structure with the complete genome sequence should allow a more accurate definition of the taxonomic position of *A. tumefaciens* in the Rhizobiaceae.

One striking finding from our analysis is the extensive similarity of the circular chromosomes of *A. tumefaciens* and the plant symbiont *S. meliloti*, which supports the view that these bacteria originated from a recent common ancestor. Galibert *et al.* speculate that the *S. meliloti* chromosome was present in a progenitor that later acquired pSymA and pSymB (*12*). The mosaic structure of the *A. tumefaciens* linear chromosome and plasmids, predominantly composed of orthologs found on each of the *S. meliloti* replicons, suggests that these organisms diverged after acquisition of the pSymA and pSymB ancestral molecules by this progenitor.

Recent models of bacterial evolution suggest that the differential acquisition and loss of genes in organisms that inhabit the same environment allows divergence into symbiotic and pathogenic lifestyles (*50*). The acquisition of such elements is apparent in both *A. tumefaciens* and *S. meliloti*. The *nod* genes of *S. meliloti* (*12, 51*), as well as the *vir* genes and T-DNA of *A. tumefaciens*, display GC content and codon usage distinct from the rest of the genome, which suggests recent evolutionary acquisition. In the case of the T-DNA, reduced GC content may facilitate expression in the plant host, where lower GC content is common. Moreover, none of the T-DNA and few of the *vir* genes of *A. tumefaciens* have orthologs in *S. meliloti*, and most *nod* genes are not found in *A. tumefaciens*. Differential selection and maintenance of such horizontally acquired genes likely led to the divergence into pathogenic and symbiotic states. Thus, these organisms provide a rich model system for further investigations into the evolutionary divergence of pathogens and symbionts.

As the central biological tool in the generation of transgenic plants for research and agriculture, *A. tumefaciens*, and the availability of its genome sequence, will continue to have an impact on plant biotechnology. Detailed studies, supplemented by this sequence, should lead to a directed refinement of plant transformation that increases both the host range and transformation efficiency of this versatile genetic tool. Genes likely to be targeted by such work include potential virulence factors that are shared between plant and animal pathogens. Examination of these genes in the genetically tractable *Agrobacterium* system may also serve to elucidate the molecular role they play in animal pathogens. It is our hope that this work will broaden the scientific foundation from which to address the worldwide debate over the production, use, and safety of genetically modified organisms.

Table 2. Number of orthologous genes of *A. tumefaciens* with respect to *S. meliloti*. The number of orthologous genes is shown in bold, with the percentage of each *A. tumefaciens* replicon they represent shown in square brackets. The remainder of the genes, which are not orthologs, are shown in the last row. Numbers of putative protein coding genes for each replicon are shown in parentheses.

		A. tumefaciens			
		Circular (2789)	Linear (1882)	pAtC58 (550)	pTiC58 (198)
S. meliloti	Chromosome (3341)	**1867** [67%]	**673** [36%]	**114** [21%]	**30** [15%]
	pSymA (1293)	**104** [4%]	**218** [12%]	**118** [21%]	**34** [17%]
	pSymB (1570)	**221** [8%]	**478** [25%]	**108** [20%]	**23** [12%]
	Nonorthologous	**597** [21%]	**513** [27%]	**210** [38%]	**111** [56%]

References and Notes

1. E. F. Smith, C. O. Townsend, *Science* **25**, 671 (1907).
2. M. DeCleene, J. DeLey, *Bot. Gaz.* **42**, 389 (1976).
3. L. Moore, in *Biology of the Rhizobiaceae*, K. L. Giles, A. G. Atherly, Eds. (Academic Press, New York, 1981), pp. 15–46.
4. J. Zupan, T. R. Muth, O. Draper, P. Zambryski, *Plant J.* **23**, 11 (2000).
5. P. J. Christie, *Mol. Microbiol.* **40**, 294 (2001).
6. N. Inon de Iannino, G. Briones, M. Tolmasky, R. A. Ugalde, *J. Bacteriol.* **180**, 4392 (1998).
7. A. Sola-Landa *et al.*, *Mol. Microbiol.* **29**, 125 (1998).
8. M. J. de Groot, P. Bundock, P. J. Hooykaas, A. G. Beijersbergen, *Nature Biotechnol.* **16**, 839 (1998).

NiceProt View of TrEMBL: O29897

Note: most headings are clickable, even if they don't appear as links. They link to the user manual or other documents.

Entry information

Entry name	**O29897**
Primary accession number	**O29897**
Secondary accession numbers	None
Entered in TrEMBL in	Release 05, January 1998
Sequence was last modified in	Release 05, January 1998
Annotations were last modified in	Release 24, June 2003

Name and origin of the protein

Protein name	**Conserved hypothetical transmembrane protein**
Synonyms	None
Gene name	**AF0350**
From	Archaeoglobus fulgidus [TaxID: 2234]
Taxonomy	Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales; Archaeoglobaceae; Archaeoglobus.

References

[1] SEQUENCE FROM NUCLEIC ACID.
STRAIN=VC-16 / DSM 4304 / ATCC 49558;
MEDLINE=98049343; PubMed=9389475; [NCBI, ExPASy, EBI, Israel, Japan]
Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E., Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D., Richardson D.L., Kerlavage A.R., Graham D.E., Kyrpides N.C., Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S., Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B., Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L., Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T., Cotton M.D., Spriggs T., Artiach P., Kaine B.P., Sykes S.M., Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A., Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
"The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon Archaeoglobus fulgidus.";
Nature 390:364-370(1997).

Comments

None

Cross-references

EMBL	AE001080; AAB90884.1; -.[EMBL / GenBank / DDBJ] [CoDingSequence]
TIGR	AF0350; -.
GO	GO:0016021;Cellular component: integral to membrane *(inferred from electronic annotation).*
ProDom	[Domain structure / List of seq. sharing at least 1 domain]
HOBACGEN	[Family / Alignment / Tree]
ProtoMap	O29897.
PRESAGE	O29897.
ModBase	O29897.
SMR	O29897; 927825BC1EFC1B46.
SWISS-2DPAGE	Get region on 2D PAGE.
UniRef	View cluster of proteins with at least 50% / 90% identity.

Keywords

Hypothetical protein; **Transmembrane**; **Complete proteome**.

Features

None

Sequence information

Length: **211** **AA** Molecular weight: **23303** **Da** CRC64: **927825BC1EFC1B46** [This is a checksum on the sequence]

```
        10         20         30         40         50         60
         |          |          |          |          |          |
MNGPDNRHAA FRYRADSRNA RRSAWHWWLR YNAPCPRIPL QLPAAGSDRD DDNGCDFNRK

        70         80         90        100        110        120
         |          |          |          |          |          |
LRSCRPHKAE ERGLLNCKGC CGKRCGWGCS WEHNLLLHRK PAVASELHPR LCLPLRQHRM

       130        140        150        160        170        180
         |          |          |          |          |          |
IYEGTIKRAM PEKPGNEVPG SSTAKATIGF VIGIITGIVG LGGGYALVPS FIYLLGSAVK

       190        200        210
         |          |          |
IAVGTSLAEV LLPPFLPQTK VHMEGNQKGS L
```

O29897 in FASTA format

View entry in original TrEMBL format
View entry in raw text format (no links)
Request for annotation of this TrEMBL entry

BLAST BLAST submission on ExPASy/SIB or at NCBI (USA)

Sequence analysis tools: ProtParam, ProtScale, Compute pI/Mw, PeptideMass, PeptideCutter, Dotlet (Java)

ScanProsite, MotifScan

Search the SWISS-MODEL Repository

ExPASy Home page | Site Map | Search ExPASy | Contact us | Swiss-Prot

Hosted by YPRC Korea | Mirror sites: | Australia | Bolivia | Canada | China | Switzerland | Taiwan | USA

10. South, T. L., Blake, P. R., Hare, D. R. & Summers, M. F. C-terminal retroviral-type zinc finger domain from the HIV-1 nucleocapsid protein is structurally similar to the N-terminal zinc finger domain. *Biochemistry* **30**, 6342–6349 (1991).
11. Luisi, B. F. *et al.* Crystallographic analysis of the interaction of the glucorticoid receptor with DNA. *Nature* **352**, 497–505 (1991).
12. Marmorstein, R., Carey, M., Ptashne, M. & Harrison, S. C. DNA recognition by GAL4: structure of a protein–DNA complex. *Nature* **356**, 408–414 (1992).
13. Everett, R. D. *et al.* A novel arrangement of zinc-binding residues and secondary structure in the C3HC4 motif of an alpha herpes virus protein family. *J. Mol. Biol.* **234**, 1038–1047 (1993).
14. Barlow, P. N., Luisi, B., Milner, A., Elliot, M. & Everett, R. Structure of the C_3HC_4 domain by ^{1}H-nuclear magnetic resonance spectroscopy. *J. Mol. Biol.* **237**, 201–211 (1994).
15. Borden, K. L. B. *et al.* The solution structure of the RING finger domain from the acute promyelocytic leukaemia proto-oncoprotein PML. *EMBO J.* **14**, 1532–1541 (1995).
16. Phillips, S. E. V. The β-ribbon DNA recognition motif. *Ann. Rev. Biophys. Biomol. Struct.* **23**, 671–701 (1994).
17. Kim, J. L., Nikolov, D. B. & Burley, S. K. Co-crystal structure of TBP recognizing the minor groove of a TATA element. *Nature* **365**, 520–527 (1993).
18. Kim, Y., Geiger, J. H., Hahn, S. & Sigler, P. B. Crystal structure of a yeast TBP–TATA-box complex. *Nature* **365**, 512–520 (1993).
19. Schumacher, M. A., Choi, K. Y., Zalkin, H. & Brennan, R. G. Crystal structure of LacI member, PurR, bound to DNA: minor groove binding by α-helices. *Science* **266**, 763–770 (1994).
20. Flick, K. E. *et al.* Crystallization and preliminary X-ray studies of I-PpoI: a nuclear, intron-encoded homing endonuclease from *Physarum polycephalum*. *Protein Sci.* **6**, 1–4 (1997).
21. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
22. Leslie, A. G. W. in *Joint CCP4 and ESF-EACMB Newsletter on Protein Crystallography* (Daresbury Laboratory, Warrington, UK, 1992).
23. CCP4 *The SERC (UK) Collaborative Computing Project No. 4, a Suite of Programs for Protein Crystallography* (Daresbury Laboratory, Warrington, UK, 1979).
24. QUANTA96 *X-ray Structure Analysis User's Reference* (Molecular Simulations, San Diego, 1996).
25. Brünger, A. *XPLOR version 3.1: A System for X-ray Crystallography and NMR* (Yale Univ. Press, New Haven, CT, 1992).
26. Laskowski, R. J., Macarthur, M. W., Moss, D. S. & Thornton, J. M. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 383–290 (1993).
27. Evans, S. V. SETOR: hardware-lighted three-dimensional solid model representations of macromolecules. *J. Mol. Graphics* **11**, 134–138 (1993).

Acknowledgements. We thank D. McHugh, K. Stephens and J. D. Heath for initial subcloning, purification and crystallization studies; R. Strong, K. Zhang and B. Scott for advice during the crystallographic analysis; and the beamline staff at the Advanced Light Source (NLBL laboratories), beamline 5.0.2, particularly T. Earnest, for assistance. B.L.S. and R.J.M. are funded for this project by the NIH. K.E.F. was supported by an NIH training grant and the American Heart Associaiton. M.S.J. was supported by an NSF fellowship and an NIH training grant.

Correspondence and requests for materials and coordinates should be addressed to B.L.S. (e-mail: bstoddar@fred.fhcrc.org). Coordinates have been deposited in the Brookhaven Protein Data Bank (accession nos lipp, 1a73, 1a74).

corrections

Emergence of symbiosis in peptide self-replication through a hypercyclic network

David H. Lee, Kay Severin, Yohei Yokobayashi & M. Reza Ghadiri

Nature **390**, 591–594 (1997)

Hypercycles are based on second-order (or higher) autocatalysis and defined by two or more replicators that are connected by another superimposed autocatalytic cycle. Our study describes a mutualistic relationship between two replicators, each catalysing the formation of the other, that are linked by a superimposed catalytic cycle. Although the kinetic data suggest the intermediary of higher-order species in the autocatalytic processes, the present system should not be referred to as an example of a minimal hypercycle in the absence of direct experimental evidence for the autocatalytic cross-coupling between replicators. □

The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon *Archaeoglobus fulgidus*

Hans-Peter Klenk, Rebecca A. Clayton, Jean-Francois Tomb, Owen White, Karen E. Nelson, Karen A. Ketchum, Robert J. Dodson, Michelle Gwinn, Erin K. Hickey, Jeremy D. Peterson, Delwood L. Richardson, Anthony R. Kerlavage, David E. Graham, Nikos C. Kyrpides, Robert D. Fleischmann, John Quackenbush, Norman H. Lee, Granger G. Sutton, Steven Gill, Ewen F. Kirkness, Brian A. Dougherty, Keith McKenney, Mark D. Adams, Brendan Loftus, Scott Peterson, Claudia I. Reich, Leslie K. McNeil, Jonathan H. Badger, Anna Glodek, Lixin Zhou, Ross Overbeek, Jeannine D. Gocayne, Janice F. Weidman, Lisa McDonald, Teresa Utterback, Matthew D. Cotton, Tracy Spriggs, Patricia Artiach, Brian P. Kaine, Sean M. Sykes, Paul W. Sadow, Kurt P. D'Andrea, Cheryl Bowman, Claire Fujii, Stacey A. Garland, Tanya M. Mason, Gary J. Olsen, Claire M. Fraser, Hamilton O. Smith, Carl R. Woese & J. Craig Venter

Nature **390**, 364–370 (1997)

The pathway for sulphate reduction is incorrect as published: in Fig. 3 on page 367, adenylyl sulphate 3-phosphotransferase (*cysC*) is not needed in the pathway as outlined, as adenylyl sulphate reductase (*aprAB*) catalyses the first step in the reduction of adenylyl sulphate. The correct sequence of reactions is: sulphate is first activated to adenylyl sulphate, then reduced to sulphite and subsequently to sulphide. The enzymes catalysing these reactions are: sulphate adenylyltransferase (*sat*), adenylylsulphate reductase (*aprAB*), and sulphite reductase (*dsrABD*). We thank Jens-Dirk Schwenn for bringing this error to our attention. □

The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon *Archaeoglobus fulgidus*

Hans-Peter Klenk*, Rebecca A. Clayton*, Jean-Francois Tomb*, Owen White*, Karen E. Nelson*, Karen A. Ketchum*, Robert J. Dodson*, Michelle Gwinn*, Erin K. Hickey*, Jeremy D. Peterson*, Deiwood L. Richardson*, Anthony R. Kerlavage*, David E. Graham†, Nikos C. Kyrpides†, Robert D. Fleischmann*, John Quackenbush*, Norman H. Lee*, Granger G. Sutton*, Steven Gill*, Ewen F. Kirkness*, Brian A. Dougherty*, Keith McKenney*, Mark D. Adams*, Brendan Loftus*, Scott Peterson*, Claudia I. Reich†, Leslie K. McNeil†, Jonathan H. Badger†, Anna Glodek*, Lixin Zhou*, Ross Overbeek‡, Jeannine D. Gocayne*, Janice F. Weidman*, Lisa McDonald*, Teresa Utterback*, Matthew D. Cotton*, Tracy Spriggs*, Patricia Artiach*, Brian P. Kaine†, Sean M. Sykes*, Paul W. Sadow*, Kurt P. D'Andrea*, Cheryl Bowman*, Claire Fujii*, Stacey A. Garland*, Tanya M. Mason*, Gary J. Olsen†, Claire M. Fraser*, Hamilton O. Smith*, Carl R. Woese† & J. Craig Venter*

* *The Institute for Genomic Research (TIGR), Rockville, Maryland 20850, USA*
† *Department of Microbiology, University of Illinois, Champaign-Urbana, Illinois 61801, USA*
‡ *Mathematics and Computer Science Division, Argonne National Laboratory, Illinois 60439, USA*

Archaeoglobus fulgidus **is the first sulphur-metabolizing organism to have its genome sequence determined. Its genome of 2,178,400 base pairs contains 2,436 open reading frames (ORFs). The information processing systems and the biosynthetic pathways for essential components (nucleotides, amino acids and cofactors) have extensive correlation with their counterparts in the archaeon *Methanococcus jannaschii*. The genomes of these two Archaea indicate dramatic differences in the way these organisms sense their environment, perform regulatory and transport functions, and gain energy. In contrast to *M. jannaschii*, *A. fulgidus* has fewer restriction-modification systems, and none of its genes appears to contain inteins. A quarter (651 ORFs) of the *A. fulgidus* genome encodes functionally uncharacterized yet conserved proteins, two-thirds of which are shared with *M. jannaschii* (428 ORFs). Another quarter of the genome encodes new proteins indicating substantial archaeal gene diversity.**

Biological sulphate reduction is part of the global sulphur cycle, ubiquitous in the earth's anaerobic environments, and is essential to the basal workings of the biosphere. Growth by sulphate reduction is restricted to relatively few groups of prokaryotes; all but one of these are Eubacteria, the exception being the archaeal sulphate reducers in the Archaeoglobales[1,2]. These organisms are unique in that they are unrelated to other sulphate reducers, and because they grow at extremely high temperatures[3]. The known Archaeoglobales are strict anaerobes, most of which are hyperthermophilic marine sulphate reducers found in hydrothermal environments[2,4] and in subsurface oil fields[5]. High-temperature sulphate reduction by *Archaeoglobus* species contributes to deep subsurface oil-well 'souring' by producing iron sulphide, which causes corrosion of iron and steel in oil- and gas-processing systems[5].

Archaeoglobus fulgidus VC-16 (refs 2, 4) is the type strain of the Archaeoglobales. Cells are irregular spheres with a glycoprotein envelope and monopolar flagella. Growth occurs between 60 and 95 °C, with optimum growth at 83 °C and a minimum division time of 4 h. The organism grows organoheterotrophically using a variety of carbon and energy sources, but can grow lithoautotrophically on hydrogen, thiosulphate and carbon dioxide[6]. We sequenced the genome of *A. fulgidus* strain VC-16 as an example of a sulphur-metabolizing organism and to gain further insight into the Archaea[7,8] through genomic comparison with *Methanococcus jannaschii*[9].

General features of the genome

The genome of *A. fulgidus* consists of a single, circular chromosome of 2,178,400 base pairs (bp) with an average of 48.5% G+C content (Fig. 1). There are three regions with low G+C content (<39%), two rich in genes encoding enzymes for lipopolysaccharide (LPS) biosynthesis, and two regions of high G+C content (>53%), containing genes for large ribosomal RNAs, proteins involved in haem biosynthesis (*hemAB*), and several transporters (Table 1). Because the origins of replication in Archaea are not characterized, we arbitrarily designated base pair one within a presumed non-coding region upstream of one of three areas containing multiple short repeat elements.

Open reading frames. Two independent coding analysis programs and BLASTX[10] searches (see Methods) predicted 2,436 ORFs (Figs 1, 2, Tables 1, 2) covering 92.2% of the genome. The average size of the *A. fulgidus* ORFs is 822 bp, similar to that of *M. jannaschii* (856 bp), but smaller than that in the completely sequenced eubacterial genomes (949 bp). All ORFs were searched against a non-redundant protein database, resulting in 1,797 putative identifications that were assigned biological roles within a classification system adapted from ref. 11. Predicted start codons are 76% ATG, 22% GTG and 2% TTG. Unlike *M. jannaschii*, where 18 inteins were found in coding regions, no inteins were identified in *A. fulgidus*. Compared with *M. jannaschii*, *A. fulgidus* contains a large number of gene duplications, contributing to its larger genome size. The average protein relative molecular mass (M_r) in *A. fulgidus* is 29,753, ranging from 1,939 to 266,571, similar to that observed in other prokaryotes. The isoelectric point (pI) of predicted proteins among sequenced prokaryotes exhibits a bimodal distribution with peaks at pIs of approximately 5.5 and 10.5. The exceptions to this are *Mycoplasma genitalium* in which the distribution is skewed towards high pI

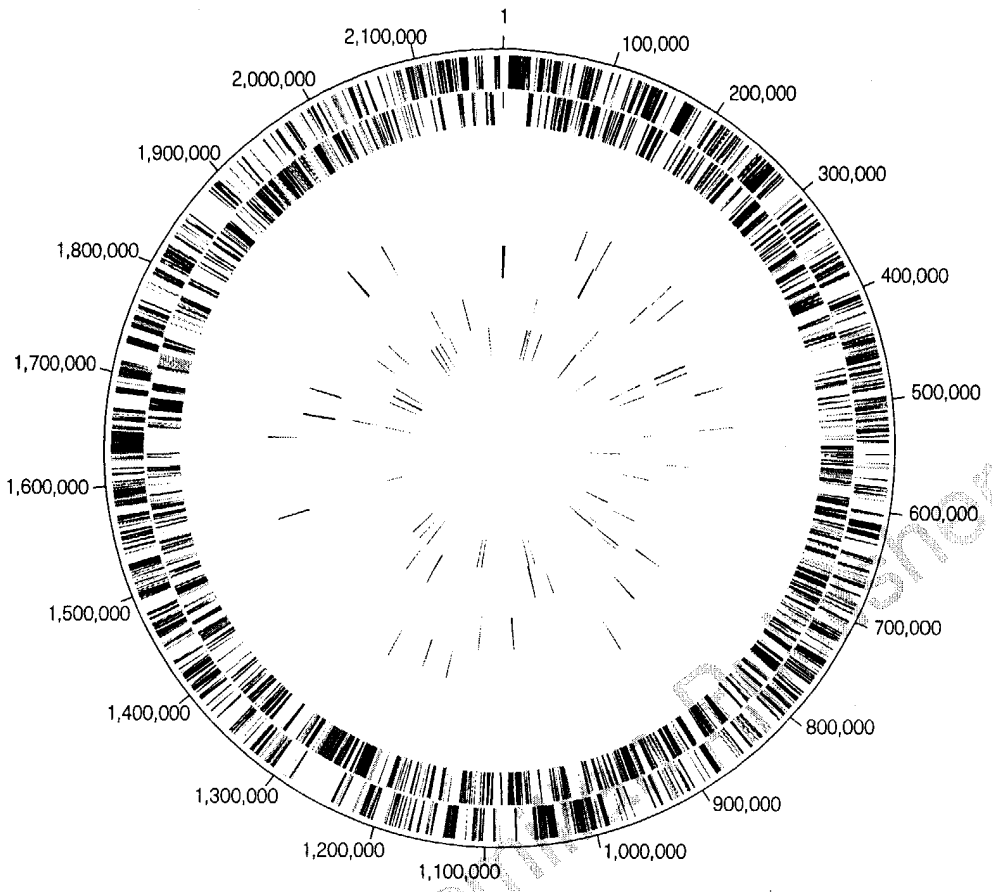


Figure 1 Circular representation of the *A. fulgidus* genome. The outer circle shows predicted protein-coding regions on the plus strand classified by function according to the colour code in Fig. 2 (except for unknowns and hypotheticals, which are in black). Second circle shows predicted protein-coding regions on the minus strand. Third and fourth circles show IS elements (red) and other repeats (green) on the plus and minus strand. Fifth and sixth circles show tRNAs (blue), rRNAs (red) and sRNAs (green) on the plus and minus strand, respectively.

Table 1 Genome features

General		
Chromosome size:	2,178,400 bp	
Protein coding regions:	92.2%	
Stable RNAs:	0.4%	
Predicted protein coding sequences:	2,436 (1.1 per kb)	
Identified by database match:	1,797	
putative function assigned:	1,096	
homologues of *M. jannaschii* ORFs:	916	
conserved hypothetical proteins:	651	
No database match:	639	
Members of 242 paralogous families:	719	
Members of 158 families with known functions:	475	
Stable RNAs	*Coordinates*	
16S rRNA:	1,790,478-1,788,987	
23S rRNA	1,788,751-1,785,820	
5S rRNA:	81,144-81,021	
7S RNA:	798,067-798,376	
RNase P:	86,281-86,032	
46 species of tRNA:	no significant clusters	
tRNAs with 15-62 bp introns:	Asp^{GUC}, Glu^{UUC}, Leu^{CAA}, Trp^{CCA}, Tyr^{GUA}	
Distinct G+C content regions	*Coordinates*	
HGC-1, >53% G+C	1,786,000-1,797,000	
HGC-2, >53% G+C	2,158,000-2,159,000	
LGC-1, <39% G+C	281,000-284,000	
LGC-2, <39% G+C	544,000-550,000	
LGC-3, <39% G+C	1,175,000-1,177,000	
Short, non-coding repeats	*Coordinates*	
SR-1A, CTTTCAATCCCATTTTGGTCTGATTTCAAC	147-4,213	
SR-1B, CTTTCAATCCCATTTTGGTCTGATTTCAAC	398,368-401,590	
SR-2, CTTTCAATCTCCATTTTCAGGGCCTCCCTTTCTTA	1,690,930-1,694,104	
Long, coding repeats	*Length*	*Copy number*
LR-01 NADH-flavin oxidoreductase	1,886 bp	2 copies
LR-02 NifS, NifU + ORF	1,549 bp	2 copies
LR-03 ISA1214 putative transposase + ISORF2	1,214 bp	6 copies
LR-04 ISA1083 putative transposase + ISORF2	1,083 bp	3 copies
LR-05 type II secretion system protein	1,014 bp	4 copies
LR-06 ISA0963 putative transposase	963 bp	7 copies
LR-07 homologue of MJ0794	836 bp	3 copies
LR-08 conserved hypothetical protein	696 bp	2 copies
LR-09 conserved hypothetical protein	628 bp	2 copies

(median, 9.8) and *A. fulgidus* where the skew is toward low pI (median, 6.3).

Multigene families. In *A. fulgidus* 719 genes (30% of the total) belong to 242 families with two or more members (Table 1). Of these families, 157 contained genes with biological roles. Most of these families contain genes assigned to the 'energy metabolism', 'transport and binding proteins', and 'fatty acid and phospholipid metabolism' categories (Table 2). The superfamily of ATP-binding subunits of ABC transporters is the largest, containing 40 members. The importance of catabolic degradation and signal recognition systems is reflected by the presence of two large superfamilies: acyl-CoA ligases and signal-transducing histidine kinases. *A. fulgidus* does not contain a homologue of the large 16-member family found in *M. jannaschii*[9].

Repetitive elements. Three regions of the *A. fulgidus* genome contain short (<40 bp) direct repeats (Table 1). Two regions (SR-1A and SR-1B) contain 48 and 60 copies, respectively, of an identical 30-bp repeat interspersed with unique sequences averaging 40 bp. The third region (SR-2) contains 42 copies of a 37-bp repeat similar in sequence to the SR-1 repeat and interspersed with unique sequence averaging 41 bp. These repeated sequences are similar to the short repeated sequences found in *M. jannaschii*.

Nine classes of long (>500 bp) repeated sequences with ≥95% sequence identity were found (LR1-LR9; Table 1). LR-3 is a novel element with 14-bp inverted repeats and two genes, one of which has weak similarity to a transposase from *Halobacterium salinarium*. One copy of LR-3 interrupts AF2090, a homologue of a large *M. jannaschii* gene encoding a protein of unknown function. LR-4 and LR-6 encode putative transposases not identified in *M. jannaschii* that may represent IS elements. The remaining LR elements are not similar to known IS elements.

Central intermediary and energy metabolism

Sulphur oxide reduction may be the dominant respiratory process in anaerobic marine and freshwater environments, and is an important aspect of the sulphur cycle in anaerobic ecosystems[12]. In this pathway, sulphate (SO_4^{2-}) is first activated to adenylylsulphate (adenosine-5′-phosphosulphate; APS), then reduced to sulphite and subsequently to sulphide[1,13] (Fig. 3). The most important enzyme in dissimilatory sulphate reduction, adenylylsulphate reductase, reduces the activated sulphate to sulphite, releasing AMP. In *A. fulgidus*, the APS reductase has a high degree of similarity and identical physiological properties to APS reductases in sulphate-reducing delta proteobacteria[14]. A desulphoviridin-type sulphite reductase then adds six electrons to sulphite to produce sulphide. As in the Eubacteria, three sulphite-reductase genes, *dsrABD*, constitute an operon. The genes for adenylylsulphate reductase and sulphate adenylyltransferase reside in a separate operon. In *A. fulgidus*, sulphate can be replaced as an electron acceptor by both thiosulphate ($S_2O_3^{2-}$) and sulphite (SO_3^{2-}), but not by elemental sulphur.

A. fulgidus VC-16 has been shown to use lactate, pyruvate, methanol, ethanol, 1-propanol and formate as carbon and energy sources[2]. Glucose has been described as a carbon source[1], but neither an uptake-transporter nor a catabolic pathway could be identified. Although it has been reported that *A. fulgidus* is incapable of growth on acetate[6], multiple genes for acetyl-CoA synthetase (which converts acetate to acetyl-CoA) were found. The organism may degrade a variety of hydrocarbons and organic acids because of the presence of 57 β-oxidation enzymes, at least one lipase, and a minimum of five types of ferredoxin-dependent oxidoreductases (Fig. 3). The predicted β-oxidation system is similar to those in Eubacteria and mitochondria, and has not previously been described in the Archaea. *Escherichia coli* requires both the *fadD* and *fadL* gene products to import long-chain fatty acids across the cell envelope into the cytosol[15]. *A. fulgidus* has 14 acyl-CoA ligases related to FadD, but as expected given that it has no outer membrane, no FadL. In *E. coli*, FadB has several metabolic functions, but in *A. fulgidus* these functions seem to be distributed among separate enzymes. For example, AF0435 encodes an orthologue of enoyl-CoA hydratase and resembles the amino-terminal domain of FadB. This gene is immediately upstream of a gene encoding an orthologue of 3-hydroxyacyl-CoA dehydrogenase that resembles the carboxy-terminal domain of FadB.

Acetyl-CoA is degraded by *A. fulgidus* through a C_1-pathway, not by the citric acid cycle or glyoxylate bypass[6,16,17]. This degradation is catalysed through the carbon monoxide dehydrogenase (CODH) pathway that consists of a five-subunit acetyl-CoA decarboxylase/synthase complex (ACDS) and five enzymes that are typically involved in methanogenesis[18]. In *A. fulgidus*, however, reverse methanogenesis occurs, resulting in CO_2 production. All of the enzymes and cofactors of methanogenesis from formylmethanofuran to N^5-methyltetrahydromethanopterin are used, but the absence of methyl-CoM reductase eliminates the possibility of methane production by conventional pathways. Production of trace amounts of methane (<0.1 μmol ml^{-1})[19] is probably a result of the reduction of N^5-methyltetrahydromethanopterin to methane and tetrahydromethanopterin by carbon monoxide (CO) dehydrogenase.

A. fulgidus also contains genes suggesting it has a second CO dehydrogenase system, homologous to that which enables *Rhodospirillum rubrum* to grow without light using CO as its sole energy source. Genes were detected for the nickel-containing CO dehydrogenase (CooS), an iron–sulphur redox protein, and a protein associated with the incorporation of nickel in CooS. These represent elements of a system that could catalyse the conversion of CO and H_2O to CO_2 and H_2.

In contrast to *M. jannaschii*, *A. fulgidus* contains genes representing multiple catabolic pathways. Systems include CoA-SH-dependent ferredoxin oxidoreductases specific for pyruvate, 2-ketoisovalerate, 2-ketoglutarate and indolepyruvate, as well as a 2-oxoacid with little substrate specificity[20,21]. Four genes with similarity to the tungsten-containing aldehyde ferredoxin oxidoreductase were also found[22].

Biochemical pathways characteristic of eubacterial metabolism, including the pentose-phosphate pathway, the Entner–Doudoroff pathway, glycolysis and gluconeogenesis, are either completely absent or only partly represented (Fig. 3). *A. fulgidus* does not have typical eubacterial polysaccharide biosynthesis machinery, yet it has been shown to produce a protein and carbohydrate-containing biofilm[23]. Nitrogen is obtained by importing inorganic molecules or degrading amino acids (Fig. 3); neither a glutamate dehydrogenase nor a relevant *fix* or *nif* gene is present.

The $F_{420}H_2$:quinone oxidoreductase complex[24] is recognized as

Figure 2 Linear representation of the *A. fulgidus* genome illustrating the location of each predicted protein-coding region, RNA gene, and repeat element in the genome. Symbols for the transporters are as follows: AsO, arsenite; COH, sugar; P_i, phosphate; aa2, dipeptide; NH_4^+, ammonium; a/o, arginine/lysine/ornithine; s/p, spermidine/putrescine; glyc, glycerol; Cl^-, chloride; Fe^{2+}, iron(II); Fe^{3+}, iron(III); I, L, V, branched-chain amino acids; P, proline; pan, pantothenate; rib, ribose; lac, lactate; Mg^{2+}/Co^{2+}, magnesium and cobalt; gln, glutamine; NO^{3-}, nitrate; ox/for, oxalate/formate; maln, malonic acid; Hg^{2+}, mercury; phs, polysaccharide; SO_4^{2-}, sulphate; OCN^-, cyanate; hex, hexuronate; phs, polysialic acid; K^+, potassium channel; H^+/Na^+, sodium/proton antiporter; Na^+/Cl^-, sodium- and chloride-dependent transporter; P/G, osmoprotection protein; Cu^{2+}, copper-transporting ATPase; +?, cation-transporting ATPase; ?, ABC-transporter without known function. Triplets associated with tRNAs represent the anticodon sequence. Numbers associated with GES represent the number of membrane-spanning domains (MSDs) according to Goldman, Engelman and Steiz scale as determined by TopPred[39]. Genes whose identification is based on genes in *M. jannaschii* are indicated by circles. Of the 236 proteins containing at least one MSD, 124 of these had two or more MSDs.

the main generator of proton-motive force. However, our analysis indicates the presence of heterodisulphide reductase and several molybdopterin-binding oxidoreductases, with polysulphide, nitrate, dimethyl sulphoxide, and thiosulphate as potential substrates, which might contribute to energizing the cell membrane. *A. fulgidus* contains a large number of flavoproteins, iron–sulphur proteins and iron-binding proteins that contribute to the general intracellular flow of electrons (Fig. 3). Detoxification enzymes include a peroxidase/catalase, an alkyl-hydroperoxide reductase, arsenate reductase, and eight NADH oxidases, presumably catalysing the

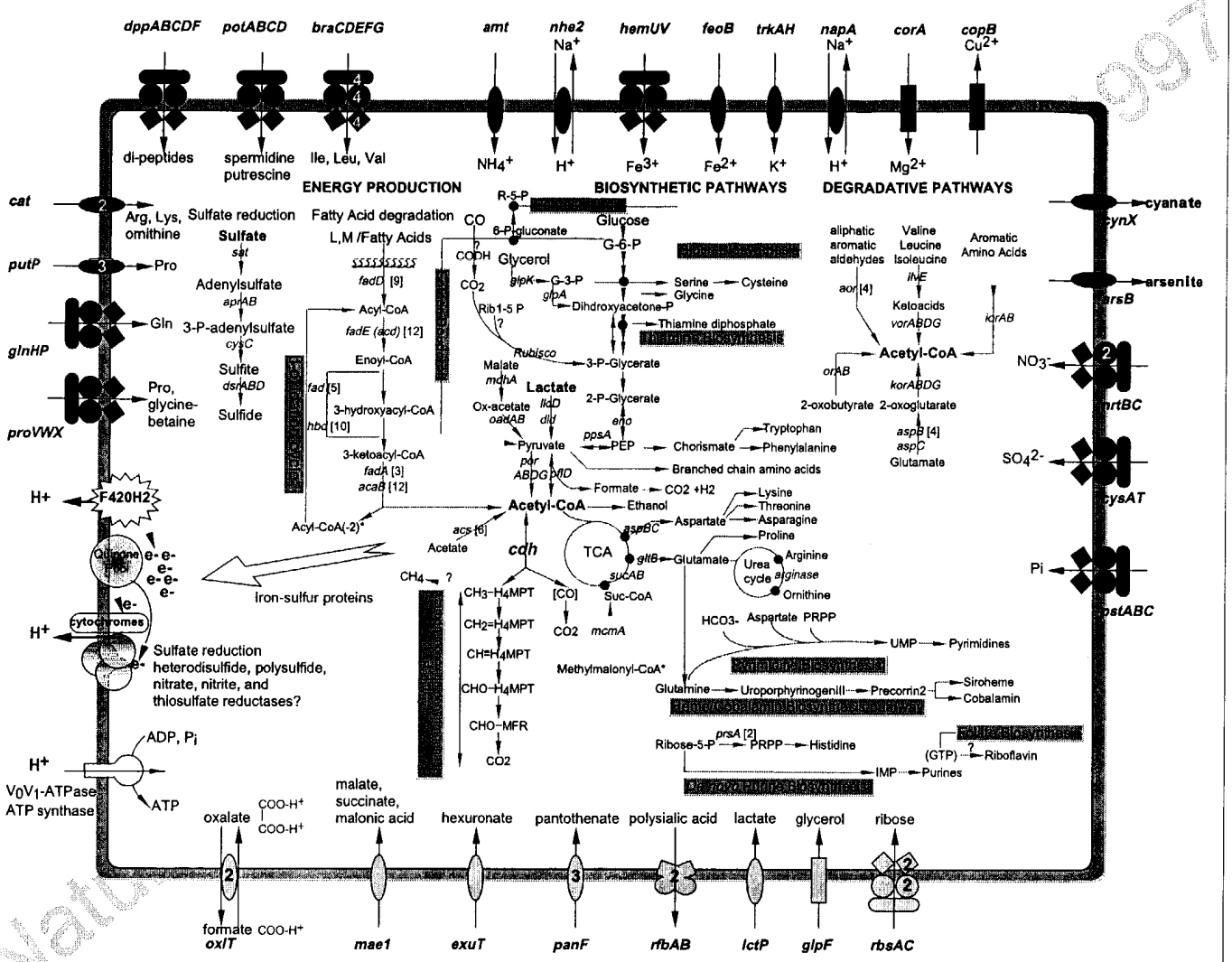


Figure 3 An integrated view of metabolism and solute transport in *A. fulgidus*. Biochemical pathways for energy production, biosynthesis of organic compounds, and degradation of amino acids, aldehydes and acids are shown with the central components of *A. fulgidus* metabolism, sulphate, lactate and acetyl-CoA highlighted. Pathways or steps for which no enzymes were identified are represented by a red arrow. A question mark is attached to pathways that could not be completely elucidated. Macromolecular biosynthesis of RNA, DNA and ether lipids have been omitted. Membrane-associated reactions that establish the proton-motive force (PMF) and generate ATP (electron transport chain and V_1V_0-ATPase) are linked to cytosolic pathways for energy production. The oxalate-formate antiporters (*oxlT*) may also contribute to the PMF by mediating electrogenic anion exchange. Each gene product with a predicted function in ion or solute transport is illustrated. Proteins are grouped by substrate specificity with transporters for cations (green), anions (red), carbohydrates/organic alcohols/acids (yellow), and amino acids/peptides/amines (blue) depicted. Ion-coupled permeases are represented by ovals (*mae1*, *exuT*, *panF*, *lctP*, *arsB*, *cynX*, *napA/nhe2*, *amt*, *feoB*, *trkAH*, *cat* and *putP* encode transporters for malate, hexuronate, pantothenate, lactate, arsenite, cyanate, sodium, ammonium, iron (II), potassium, arginine/lysine and proline, respectively). ATP-binding cassette (ABC) transport systems are shown as composite figures of ovals, diamonds and circles (*proVWX*, *glnHPQ*, *dppABCDF*, *potABCD*, *braCDEFG*, *hemUV*, *nrtBC*, *cysAT*, *pstABC*, *rbsAC*, *rfbAB* correspond to gene products for proline, glutamine, dipeptide, spermidine/putrescine, branch-chain amino acids, iron (III), nitrate, sulphate, phosphate, ribose and polysialic acid transport, respectively). All other porters drawn as rectangles (*glpF*, glycerol uptake facilitator; *copB*, copper transporting ATPase; *corA*, magnesium and cobalt transporter). Export and import of solutes is designated by arrows. The number of paralogous genes encoding each protein is indicated in brackets for cytoplasmic enzymes, or within the figure for transporters. Abbreviations: *acs*, acetyl-CoA synthetase; *aor*, aldehyde ferredoxin oxidoreductase; *aprAB*, adenylylsulphate reductase; *aspBC*, aspartate aminotransferase; *cdh*, acetyl-CoA decarbonylase/synthase complex; *cysC*, adenylylsulphate 3-phosphotransferase; *dld*, D-lactate dehydrogenase; *dsrABD*, sulphite reductase; *eno*, enolase; *fadA/acaB*, 3-ketoacyl-CoA thiolase; *fadD*, long-chain-fatty-acid-CoA ligase; *fad*, enoyl-CoA hydratase; *fadE* (*acd*), acyl-CoA dehydrogenase; *glpA*, glycerol-3-phosphate dehydrogenase; *glpK*, glycerol kinase; *gltB*, glutamate synthase; *hbd*, 3-hydroxyacyl-CoA dehydrogenase; *ilvE*, branched-chain amino-acid aminotransferase; *iorAB*, indolepyruvate ferredoxin oxidoreductase; *korABDG*, 2-ketoglutarate ferredoxin oxidoreductase; *lldD*, L-lactate dehydrogenase; *mcmA*, methylmalonyl-CoA mutase; *mdhA*, L-malate dehydrogenase; *oadAB*, oxaloacetate decarboxylase; *orAB*, 2-oxoacid ferredoxin oxidoreductase; *pflD*, pyruvate formate lysase 2; *porABDG*, pyruvate ferredoxin oxidoreductase; *ppsA*, phosphoenolpyruvate synthase; *prsA*, ribose-phosphate pyrophosphokinase; *sucAB*, 2-ketoglutarate dehydrogenase; *sat*, sulphate adenylyltransferase; TCA, tricarboxylic acid cycle; *vorABDG*, 2-ketoisovalerate ferredoxin oxidoreductase.

four-electron reduction of molecular oxygen to water, with the concurrent regeneration of NAD.

Transporters

A. fulgidus may synthesize several transporters for the import of carbon-containing compounds, probably contributing to its ability to switch from autotrophic to heterotrophic growth[5]. Both *M. jannaschii* and *A. fulgidus* have branched-chain amino-acid ABC transport systems and a transporter for the uptake of arginine and lysine. *A. fulgidus* encodes proteins for dipeptide, spermidine/putrescine, proline/glycine-betaine and glutamine uptake, as well as transporters for sugars and acids, rather like the membrane systems described in eubacterial heterotrophs. These compounds provide the necessary substrates for numerous biosynthetic and degradative pathways (Fig. 3).

Many *A. fulgidus* redox proteins are predicted to require iron. Correspondingly, iron transporters have been identified for the import of both oxidized (Fe^{3+}) and reduced (Fe^{2+}) forms of iron. There are duplications in functional and regulatory genes in both systems. The uptake of Fe^{3+} may depend on haemin or a haemin-like compound because *A. fulgidus* has orthologues to the eubacterial *hem* transport system proteins, HemU and HemV. *A. fulgidus* may also use the regulatory protein Fur to modulate Fe^{3+} transport; this protein is not present in *M. jannaschii*. Fe^{2+} uptake occurs through a modified Feo system containing FeoB. This is the third example of an isolated *feoB* gene: *M. jannaschii* and *Helicobacter pylori* also appear to lack *feoA*, implying that FeoA is not essential for iron transport in these organisms.

A complex suite of proteins regulates ionic homeostasis. Ten distinct transporters facilitate the flux of the physiological ions K^+, Na^+, NH_4^+, Mg^{2+}, Fe^{2+}, Fe^{3+}, NO_3^-, SO_4^{2-} and inorganic phosphate (P_i). Most of these transporters have homologues in *M. jannaschii* and are therefore likely to be critical for nutrient acquisition during autotrophic growth. *A. fulgidus* has additional ion transporters for the elimination of toxic compounds including copper, cyanate and arsenite. As in *M. jannaschii*, the *A. fulgidus* genome contains two paralogous operons of cobalamin biosynthesis-cobalt transporters, *cbiMQO*.

Sensory functions and regulation of gene expression

Consistent with its extensive energy-producing metabolism and versatile system for carbon utilization, *A. fulgidus* has complex sensory and regulatory networks. These networks contain over 55 proteins with presumed regulatory functions, including members of the ArsR, AsnC and Sir2 families, as well as several iron-dependent repressor proteins. There are at least 15 signal-transducing histidine kinases, but only nine response regulators; this difference suggests there is a high degree of cross-talk between kinases and regulators. Only four response regulators appear to be in operons with histidine kinases, including those in the methyl-directed chemotaxis system (Che), which lies adjacent to the flagellar biosynthesis operon. Although rich in regulatory proteins, *A. fulgidus* apparently lacks regulators for response to amino-acid and carbon starvation as well as to DNA damage. Finally, *A. fulgidus* contains a homologue of the mammalian mitochondrial benzodiazepine receptor, which functions as a sensor in signal-transduction pathways[25]. These receptors have been previously identified only in Proteobacteria and Cyanobacteria[25].

Replication, repair and cell division

A. fulgidus possesses two family B DNA polymerases, both related to the catalytic subunit of the eukaryal delta polymerase, as previously observed in the Sulfolobales[26]. It also has a homologue of the proofreading ε subunit of *E. coli* Pol III, not previously observed in the Archaea. The DNA repair system is more extensive than that found in *M. jannaschii*, including a homologue of the eukaryal Rad25, a 3-methyladenine DNA glycosylase, and exodeoxynuclease III. As well as reverse gyrase, topoisomerase I (ref. 9), and topoisomerase VI (ref. 27), the genes for the first archaeal DNA gyrase were identified.

A. fulgidus lacks a recognizable type II restriction-modification system, but contains one type I system. In contrast, two type II and three type I systems were identified in *M. jannaschii*. No homologue of the *M. jannaschii* thermonuclease was identified.

The cell-division machinery is similar to that of *M. jannaschii*, with orthologues of eubacterial *fts* and eukaryal *cdc* genes. However, several *cdc* genes found in *M. jannaschii*, including homologues of *cdc23*, *cdc27*, *cdc47* and *cdc54*, appear to be absent in *A. fulgidus*.

Transcription and translation

A. fulgidus and *M. jannaschii* have transcriptional and translational systems distinct from their eubacterial and eukaryal counterparts. In both, the RNA polymerase contains the large universal subunits and five smaller subunits found in both Archaea and eukaryotes. Transcription initiation is a simplified version of the eukaryotic mechanism[28,29]. However, *A. fulgidus* alone has a homologue of eukaryotic TBP-interacting protein 49 not seen in *M. jannaschii*, but apparently present in *Sulfolobus solfactaricus*.

Translation in *A. fulgidus* parallels *M. jannaschii* with a few exceptions. The organism has only one rRNA operon with an Ala-tRNA gene in the spacer and lacks a contiguous 5S rRNA gene. Genes for 46 tRNAs were identified, five of which contain introns in the anticodon region that are presumably removed by the intron excision enzyme EndA. The gene for selenocysteine tRNA (SelC) was not found, nor were the genes for SelA, SelB and SelD. With the exception of Asp-$tRNA^{GTC}$ and Val-$tRNA^{CAC}$, tRNA genes are not linked in the *A. fulgidus* genome. The RNA component of the tRNA maturation enzyme RNase P is present. Both *A. fulgidus* and *M. jannaschii* appear to possess an enzyme that inserts the tRNA-modified nucleoside archaeosine, but only *A. fulgidus* has the related enzyme that inserts the modified base queuine.

Both *A. fulgidus* and *M. jannaschii* lack glutamine synthetase and asparagine synthetase; the relevant tRNAs are presumably amino-acylated with glutamic and aspartic acids, respectively. An enzymatic *in situ* transamidation then converts the amino acid to its amide form, as seen in other Archaea and in Gram-positive Eubacteria[30]. Indeed, genes for the three subunits of the Glu-tRNA amidotransferase (*gatABC*) have been identified in *A. fulgidus*. The Lys aminoacyl-tRNA synthetase in both organisms is a class I-type, not a class II-type[31]. *A. fulgidus* possesses a normal tRNA synthetase for both Cys and Ser, unlike *M. jannaschii* in which the former was not identifiable and the latter was unusual[9].

M. jannaschii has a single gene belonging to the TCP-1 chaperonin family, whereas *A. fulgidus* has two that encode subunits α and β of the thermosome. Phylogenetic analysis of the archaeal TCP-1 family indicates that these *A. fulgidus* genes arose by a recent species-specific gene duplication, as is the case for the two subunits of the *Thermoplasma acidophilum* thermosome[32] and the *Sulfolobus shibatae* rosettasome[33]. As in *M. jannaschii*, no *dnaK* gene was identified.

Biosynthesis of essential components

Like most autotrophic microorganisms, *A. fulgidus* is able to synthesize many essential compounds, including amino acids, cofactors, carriers, purines and pyrimidines. Many of these biosynthetic pathways show a high degree of conservation between *A. fulgidus* and *M. jannaschii*. These two Archaea are similar in their biosynthetic pathways for siroheme, cobalamin, molybdopterin, riboflavin, thiamin and nictotinate, the role category with greatest conservation between these two organisms being amino-acid biosynthesis. Of 78 *A. fulgidus* genes assigned to amino-acid biosynthetic pathways, at least 73 (94%) have homologues in *M. jannaschii*. For both archaeal species, amino-acid biosynthetic pathways resemble those of *Bacillus subtilis* more closely than

those of *E. coli*. For example, in *A. fulgidus* and *M. jannaschii*, tryptophan biosynthesis is accomplished by seven enzymes, TrpA, B, C, D, E, F, G as in *B. subtilis*, rather than by five enzymes, TrpA, B, C, D, E (including the bifunctional TrpC and TrpD) as found in *E. coli*.

No biotin biosynthetic genes were identified, yet biotin can be detected in *A. fulgidus* cell extracts[34], and several genes encode a biotin-binding consensus sequence. Similarly, *A. fulgidus* lacks the genes for pyridoxine biosynthesis although pyridoxine can be found in cell extracts (albeit at lower levels than seen in *E. coli* and several Archaea[34]). No gene encoding ferrochelatase, the terminal enzyme in haem biosynthesis, has been identified, although *A. fulgidus* is known to use cytochromes[34]. These cofactors may be obtained by mechanisms that we have not recognized. Although all of the enzymes required for pyrimidine biosynthesis appear to be present, three enzymes in the purine pathway (GAR transformylase, AICAR formyltransferase and the ATPase subunit of AIR carboxylase) have not been identified, presumably because they exist as new isoforms.

The Archaea share a unique cell membrane composed of ether lipids containing a glycerophosphate backbone with a 2,3-*sn* stereochemistry[35] for which there are multiple biosynthetic pathways[36]. In the case of *Halobacterium cutirubrum*, the backbone is apparently obtained by enantiomeric inversion of *sn*-glycerol-3-phosphate; in *Sulfolobus acidocaldarius* and *Methanobacterium thermoautotrophicum*, *sn*-glycerol-1-phosphate dehydrogenase builds the backbone from dihydroxyacetonephosphate. An orthologue of *sn*-glycerol-1-phosphate dehydrogenase has been identified in *A. fulgidus*, suggesting that the latter pathway is present.

Conclusions

Although *A. fulgidus* has been studied since its discovery ten years ago[1], the completed genome sequence provides a wealth of new information about how this unusual organism exploits its environment. For example, its ability to reduce sulphur oxides has been well characterized, but genome sequence data demonstrate that *A. fulgidus* has a great diversity of electron transport systems, some of unknown specificity. Similarly, *A. fulgidus* has been characterized as a scavenger with numerous potential carbon sources, and its gene complement reveals the extent of this capability. *A. fulgidus* appears to obtain carbon from fatty acids through β-oxidation, from degradation of amino acids, aldehydes and organic acids, and perhaps from CO.

A. fulgidus has extensive gene duplication in comparison with other fully sequenced prokaryotes. For example, in the fatty acid and phospholipid metabolism category, there are 10 copies of 3-hydroxyacyl-CoA dehydrogenase, 12 copies of 3-ketoacyl-CoA thiolase, and 12 of acyl-CoA dehydrogenase. The duplicated proteins are not identical, and their presence suggests considerable metabolic differentiation, particularly with respect to the pathways for decomposing and recycling carbon by scavenging fatty acids. Other categories show similar, albeit less dramatic, gene redundancy. For example, there are six copies of acetyl-CoA synthetase and four aldehyde ferredoxin oxidoreductases for fermentation, as well as four copies of aspartate aminotransferase for amino-acid biosynthesis. These observations, together with the large number of paralogous gene families, suggest that gene duplication has been an important evolutionary mechanism for increasing physiological diversity in the Archaeoglobales.

A comparison of two archaeal genomes is inadequate to assess the diversity of the entire domain. Given this caveat, it is nevertheless possible to draw some preliminary conclusions from the comparison of *M. jannaschii* and *A. fulgidus*. A comparison of the gene content of these Archaea reveals that gene conservation varies significantly between role categories, with genes involved in transcription, translation and replication highly conserved; approximately 80% of the *A. fulgidus* genes in these categories have homologues in *M. jannaschii*. Biosynthetic pathways are also highly conserved, with approximately 80% of the *A. fulgidus* biosynthetic genes having homologues in *M. jannaschii*. In contrast, only 35% of the *A. fulgidus* central intermediary metabolism genes have homologues, reflecting their minimal metabolic overlap.

Over half of the *A. fulgidus* ORFs (1,290) have no assigned biological role. Of these, 639 have no database match. The remaining 651, designated 'conserved hypothetical proteins', have sequence similarity to hypothetical proteins in other organisms, two-thirds with apparent homologues in *M. jannaschii*. These shared hypothetical proteins will probably add to our understanding of the genetic repertoire of the Archaea. Analysis of the *A. fulgidus* and other archaeal and eubacterial genomes will provide the information necessary to begin to define a core set of archaeal genes, as well as to better understand prokaryotic diversity. □

Methods

Whole-genome random sequencing procedure. The type strain, *A. fulgidus* VC-16, was grown from a culture derived from a single cell isolated by optical tweezers[37] and provided by K. O. Stetter (University of Regensburg). Cloning, sequencing and assembly were essentially as described previously for genomes sequenced by TIGR[9,38–40]. One small-insert and one medium-insert plasmid library were generated by random mechanical shearing of genomic DNA. One large-insert lambda (λ) library was generated by partial *Tsp*509I digestion and ligation to λ-DASHII/*Eco*RI vector (Stratagene). In the initial random sequencing phase, 6.7-fold sequence coverage was achieved with 27,150 sequences from plasmid clones (average read length 500 bases) and 1,850 sequences from λ-clones. Both plasmid and λ-sequences were jointly assembled using TIGR assembler[41], resulting in 152 contigs separated by sequence gaps and five groups of contigs separated by physical gaps. Sequences from both ends of 560 λ-clones served as a genome scaffold, verifying the orientation, order and integrity and the contigs. Sequence gaps were closed by editing the ends of sequence traces and/or primer walking on plasmid or λ-clones clones spanning the respective gap. Physical gaps were closed by combinatorial polymerase chain reaction (PCR) followed by sequencing of the PCR product. At the end of gap closure, 90 regions representing 0.33% of the genome had only single-sequence coverage. These regions were confirmed with terminator reactions to ensure a minimum of 2-fold sequence coverage for the whole genome. The final genome sequence is based on 29,642 sequences, with a 6.8-fold sequence coverage. The linkage between the terminal sequences of 2,101 clones from the small-insert plasmid library (average size 1,419 bp) and 8,726 clones from the medium-insert plasmid library (average size 2,954 bp) supported the genome scaffold formed by the λ-clones (average size 16,381 bp), with 96.9% of the genome covered by λ-clones. The reported sequence differs in 20 positions from the 14,389 bp of DNA in a total of 11 previously published *A. fulgidus* genes.

ORF prediction and gene family identification. Coding regions (ORFs) were identified using a combination strategy based on two programs. Initial sets of ORFs were derived with GeneSmith (H.O.S., unpublished), a program that evaluates ORF length, separation and overlap between ORFs, and with CRITICA (J.H.B. & G.J.O., unpublished), a coding region identification tool using comparative analysis. The two largely overlapping sets of ORFs were merged into one joint set containing all members of both initial sets. ORFs were searched against a non-redundant protein database using BLASTX[10] and those shorter than 30 codons 'coding' for proteins without a database match were eliminated. Frameshifts were detected and corrected where appropriate as described previously[40]. Remaining frameshifts are considered authentic and corresponding regions were annotated as 'authentic frameshift'. In total, 527 hidden Markov models, based upon conserved protein families (PFAM version 2.0), were searched with HMMER to determine ORF membership in families and superfamilies[42]. Families of paralogous genes were constructed as described previously[40]. TopPred[43] was used to identify membrane-spanning domains in proteins.

Received 9 September; accepted 4 November 1997.

1. Stetter, K. O., Lauerer, G., Thomm, M. & Neuner, A. Isolation of extremely thermophilic sulfate reducers: Evidence for a novel branch of archaebacteria. *Science* **236**, 822–824 (1987).
2. Stetter, K. O., in *The Prokaryotes* (eds Balows, A., Trüper, H. G., Dworkin, M., Harder, W. & Schleifer, K. H.) 707–711 (Springer, Berlin, 1992).
3. Stetter, K. O. Microbial life in hyperthermal environments: Microorganisms from exotic environments continue to provide surprises about life's extremities. *ASM News* **61**, 285–290 (1995).

4. Stetter, K. O. *Archaeoglobus fulgidus* gen. nov., sp. nov.: a new taxon of extremely thermophilic archaebacteria. *Syst. Appl. Microbiol.* **10,** 172–173 (1988).
5. Stetter, K. O. *et al.* Hyperthermophilic archaea are thriving in deep North Sea and Alaskan oil reservoirs. *Nature* **365,** 743–745 (1993).
6. Vorholt, J., Kunow, J., Stetter, K. O. & Thauer, R. K. Enzymes and coenzymes of the carbon monoxide dehydrogenase pathway for autotrophic CO_2 fixation in *Archaeoglobus lithotrophicus* and the lack of carbon monoxide dehydrogenase in the heterotrophic *A. profundus*. *Arch. Microbiol.* **163,** 112–118 (1995).
7. Woese, C. R. & Fox, G. E. Phylogenetic structure of the prokaryotic domain: The primary kingdoms. *Proc. Natl Acad. Sci. USA* **74,** 5088–5090 (1977).
8. Woese, C. R., Kandler, O. & Wheelis, M. L. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc. Natl Acad. Sci. USA* **87,** 4576–4579 (1990).
9. Bult, C. J. *et al.* Complete genome sequence of the methanogenic archaeon *Methanococcus jannaschii*. *Science* **273,** 1058–1073 (1996).
10. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215,** 403–410 (1990).
11. Riley, M. Functions of gene products of *Escherichia coli*. *Microbiol. Rev.* **57,** 862–952 (1993).
12. Cooling, F. B. III, Maloney, C. L., Nagel, E., Tabinowski, J. & Odom, J. M. Inhibition of sulfate respiration by 1,8-dehydroxyanthraquinone and other anthraquinone derivatives. *Appl. Environ. Microbiol.* **62,** 2999–3004 (1996).
13. Thauer, R. K. & Kunow, J. in *Sulfate Reducing Bacteria* (ed. Barton, L. L.) 33–48 (Plenum, New York, 1995).
14. Speich, D. *et al.* Adenylylsulfate reductase from the sulfate-reducing archaeon *Archaeoglobus fulgidus*: cloning and characterization of the genes and comparison of the enzyme with other iron-sulfur flavoproteins. *Microbiology* **140,** 1273–1284 (1994).
15. Clark, D. P. & Cronan, J. E. Jr in Escherichia coli *and* Salmonella typhimurium: *Cellular and Molecular biology* (ed Neidhardt, F. C.) 343–357 (ASM Press, Washington DC, 1996).
16. Möller-zirkhan, D. & Thauer, R. K. Anaerobic lactate oxidation to 3 CO_2 by *Archaeoglobus fulgidus via* the carbon monoxide dehydrogenase pathway: demonstration of the acetyl-CoA carbon-carbon cleavage reaction in cell extracts. *Arch. Microbiol.* **153,** 215–218 (1990).
17. Schauder, R., Eikmanns, B., Thauer, R. K., Widdel, F. & Fuchs, G. Acetate oxidation to CO_2 in anaerobic-bacteria *via* a novel pathway not involving reactions of the citric-acid cycle. *Arch. Microbiol.* **145,** 162–172 (1986).
18. Dai, Y.-R. *et al.* Acetyl-CoA decarbonylase/synthase complex from *Archaeoglobus fulgidus*: purification, characterization, and properties. *Arch. Microbiol.* (submitted).
19. Gorris, L. G. M., Voet, A. C. W. A. & van der Drift, C. Structural characteristics of methanogenic cofactors in the non-methanogenic archaebacterium *Archaeoglobus fulgidus*. *BioFactors* **3,** 29–35 (1991).
20. Zhang, Q., Iwasaki, T., Wakagi, T. & Oshima, T. 2-oxoacid:ferredoxin oxidoreductase from the thermoacidophilic archaeon, *Sulfolobus* sp. strain 7. *J. Biochem.* **120,** 587–599 (1996).
21. Tersteegen, A., Linder, D., Thauer, R. K. & Hedderich, R. Structures and functions of four anabolic 2-oxoacid oxidoreductases in *Methanobacterium thermoautotrophicum*. *Eur. J. Biochem.* **244,** 862–868 (1997).
22. Kletzin, A. & Adams, M. W. W. Molecular and phylogenetic characterization of pyruvate and 2-ketoisovalerate ferredoxin oxidoreductases from *Pyrococcus furiosus* and pyruvate ferredoxin oxidoreductase from *Thermotoga maritima*. *J. Bacteriol.* **178,** 248–257 (1996).
23. LaPaglia, C. & Hartzell, P. L. Stress-induced production of biofilm in the hyperthermophile *Archaeoglobus fulgidus*. *Appl. Environ. Microbiol.* **63,** 3158–3163 (1997).
24. Kunow, J., Linder, D., Stetter, K. O. & Thauer, R. K. $F_{420}H_2$: quinone oxidoreductase from *Archaeoglobus fulgidus*—characterization of a membrane-bound mutlisubunit complex containing FAD and iron–sulfur clusters. *Eur. J. Biochem.* **223,** 503–511 (1994).
25. Yeliseev, A. A., Krueger, K. E. & Kaplan, S. A mammalian mitochondrial drug receptor functions as a bacterial "oxygen" sensor. *Proc. Natl Acad. Sci. USA* **94,** 5101–5106 (1997).
26. Edgell, D. R., Klenk, H.-P. & Doolittle, W. F. Gene duplications in evolution of archaeal family B DNA polymerases. *J. Bacteriol.* **179,** 2632–2640 (1997).
27. Bergerat, A. *et al.* An atypical topoisomerase II from archaea with implications for meiotic recombination. *Nature* **386,** 414–417 (1997).
28. Marsh, T. L., Reich, C. I., Whitelock, R. B. & Olsen, G. J. Transcription factor IID in the Archaea: sequences in the *Thermococcus celer* genome would encode a product closely related to the TATA-binding protein of eukaryotes. *Proc. Natl Acad. Sci. USA* **91,** 4180–4184 (1994).
29. Kosa, P. F., Ghosh, G., DeDecker, B. S. & Sigler, P. B. The 2.1-A crystal structure of an archaeal preinitiation complex: TATA-box-binding protein/transcription factor (II)B core/TATA-box. *Proc. Natl Acad. USA* **94,** 6042–6047 (1997).
30. Curnow, A. W. *et al.* Glu-$tRNA^{Gln}$ amidotransferase: a novel heterotrimeric enzyme required for correct decoding of glutamine codons during translation. *Proc. Natl Acad. Sci. USA* **94,** 11819–11826 (1997).
31. Ibba, M., Bobo, J. L., Rosa, P. A. & Soll, D. Archaeal-type lysyl-tRNA synthetase in the Lyme disease spirochete *Borrelia burgdorferi*. *Proc. Natl Acad. Sci. USA* (submitted).
32. Waldmann, T., Lupas, A., Kellermann, J., Peters, J. & Baumeister, W. Primary structure of the thermosome from *Thermoplasma acidophilum*. *Hoppe-Seyler's Biol. Chem.* **376,** 119–126 (1995).
33. Kagawa, H. K. *et al.* The 60 kDa heat shock proteins in the hyperthermophilic archaeon *Sulfolobus shibatae*. *J. Mol. Biol.* **253,** 712–725 (1995).
34. Noll, K. M. & Barber, T. S. Vitamin contents of archaebacteria. *J. Bacteriol.* **170,** 4315–4321 (1988).
35. Thornebene, T. G. & Langworthy, T. A. Diphytanyl and dibiphytanyl glycerol ether lipids of methanogenic archaebacteria. *Science* **203,** 51–53 (1979).
36. Nishihara, M. & Koga, Y. sn-glycerol-1-phosphate dehydrogenase in *Methanobacterium thermoautotrophicum*: key enzyme in biosynthesis of the enantiomeric glycerophosphate backbone of ether phospholipids of archaebacteria. *J. Biochem.* **117,** 933–935 (1995).
37. Huber, R. *et al.* Isolation of a hyperthermophilic archaeum predicted by *in situ* RNA analysis. *Nature* **376,** 57–58 (1995).
38. Fleischmann, R. D. *et al.* Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* **269,** 496–511 (1995).
39. Fraser, C. M. *et al.* The minimal gene complement of *Mycoplasma genitalium*. *Science* **270,** 397–403 (1995).
40. Tomb, J.-F. *et al.* The complete genome sequence of the gastric pathogen *Helicobacter pylori*. *Nature* **388,** 539–547 (1997).
41. Sutton, G. G., White, O., Adams, M. D. & Kerlavage, A. R. TIGR Assembler: A new tool for assembling large shotgun sequencing projects. *Genome Sequence Technol.* **1,** 9–19 (1995).
42. Sonnhammer, E. L., Eddy, S. R. & Durbin, R. Pfam: A comprehensive database of protein families based on seed alignments. *Proteins* **28,** 405–420 (1997).
43. Claros, M. G. & von Heijne, G. TopPred II: an improved software for membrane protein structure predictions. *Comput. Appl. Biosci.* **10,** 685–686 (1994).

Acknowledgements. We thank M. Heaney, J. Scott and R. Shirley for software and database support; V. Sapiro, B. Vincent, J. Meehan and D. Maas for computer system support; B. Cameron and D. J. Doyle for editorial assistance: and K. O. Stetter for providing *A. fulgidus* VC-16. This work was supported by the US Department of Energy.

Correspondence and requests for materials should be addressed to J.C.V. (e-mail: gaf@tigr.org). The annotated genome sequence and the gene family alignments are available on the World-Wide Web at http://www.tigr.org/tdb/mdb/afdb/afdb.html. The sequence has been deposited in GenBank with accession number AE000782.

Table 2. List of *A. fulgidus* genes with putative identification. Gene numbers correspond to those in Fig. 2. Percentages represent per cent identities.

AMINO ACID BIOSYNTHESIS

General

AF0906	hydantoin utilization protein A (hyuA)	27.4%

Aromatic amino acid family

AF0228	3-dehydroquinate dehydratase (aroD)	36.8%
AF1497	5-enolpyruvylshikimate 3-phosphate synthase (aroA)	41.5%
AF1603	anthranilate synthase component I (trpE)	43.7%
AF1604	anthranilate synthase component II (trpD)	43.8%
AF1602	anthranilate synthase component II (trpG)	50.0%
AF0227	chorismate mutase/prephenate dehydratase (pheA)	32.2%
AF0670	chorismate synthase (aroC)	55.3%
AF1601	phosphoribosyl anthranilate isomerase (trpF)	37.1%
AF2327	shikimate 5-dehydrogenase (aroE)	43.1%
AF0343	tryptophan repressor binding protein (wrbA)	46.6%
AF1599	tryptophan synthase, subunit alpha (trpA)	39.5%
AF1240	tryptophan synthase, subunit beta (trpB-1)	39.4%
AF1600	tryptophan synthase, subunit beta (trpB-2)	64.1%

Aspartate family

AF2112	5-methyltetrahydropteroyltriglutamate-homocysteine methyltransferase (metE)	28.1%
AF0882	asparaginase (asnA)	45.9%
AF1439	asparagine synthetase (asnB)	36.9%
AF2366	aspartate aminotransferase (aspB-1)	42.3%
AF2129	aspartate aminotransferase (aspB-2)	45.4%
AF1623	aspartate aminotransferase (aspB-3)	39.4%
AF0409	aspartate aminotransferase (aspB-4)	45.2%
AF1417	aspartate aminotransferase (aspC)	46.2%
AF0700	aspartate kinase (lysC)	49.1%
AF1422	aspartate racemase	48.0%
AF1506	aspartate-semialdehyde dehydrogenase (asd)	60.9%
AF0800	diaminopimelate decarboxylase (lysA)	45.6%
AF0747	diaminopimelate epimerase (dapF)	45.8%
AF0909	dihydrodipicolinate reductase (dapB)	48.6%
AF0910	dihydrodipicolinate synthase (dapA)	51.0%
AF0935	homoserine dehydrogenase (hom)	47.9%
AF0886	S-adenosylhomocysteinase hydrolase (ahcY-1)	31.7%
AF2000	S-adenosylhomocysteinase hydrolase (ahcY-2)	67.3%
AF0051	succinyl-diaminopimelate desuccinylase (dapE-1)	30.5%
AF0904	succinyl-diaminopimelate desuccinylase (dapE-2)	43.8%
AF0651	threonine synthase (thrC-1)	40.5%
AF1316	threonine synthase (thrC-2)	61.0%

Glutamate family

AF1280	acetylglutamate kinase (argB)	56.1%
AF2288	acetylglutamate kinase, putative	29.0%
AF0080	acetylornithine aminotransferase (argD-1)	48.3%
AF1815	acetylornithine aminotransferase (argD-2)	36.2%
AF0522	acetylornithine deacetylase (argE)	29.4%
AF0883	argininosuccinate lyase (argH)	42.2%
AF2252	argininosuccinate synthetase (argG)	62.0%
AF1147	glutamate N-acetyltransferase (argI)	47.8%
AF0953	glutamate synthase (gltB)	57.9%
AF0949	glutamine synthetase (glnA)	43.3%
AF2071	N-acetyl-gamma-glutamyl-phosphate reductase (argC)	53.3%
AF1256	ornithine carbamoyltransferase (argF)	51.7%

Pyruvate family

AF0957	2-isopropylmalate synthase (leuA-1)	53.5%
AF0219	2-isopropylmalate synthase (leuA-2)	53.9%
AF2199	3-isopropylmalate dehydratase, large subunit (leuC)	49.3%
AF0629	3-isopropylmalate dehydratase, small subunit (leuD-1)	56.4%
AF1761	3-isopropylmalate dehydratase, small subunit (leuD-2)	57.1%
AF0628	3-isopropylmalate dehydrogenase (leuB)	59.2%
AF1720	acetolactate synthase, large subunit (ilvB-1)	57.5%
AF1780	acetolactate synthase, large subunit (ilvB-2)	32.1%
AF2015	acetolactate synthase, large subunit (ilvB-3)	34.1%
AF2100	acetolactate synthase, large subunit (ilvB-4)	38.4%
AF1719	acetolactate synthase, small subunit (ilvN)	60.4%
AF1672	acetolactate synthase, small subunit, putative	29.7%
AF0933	branched-chain amino acid aminotransferase (ilvE)	59.0%
AF1014	dihydroxy-acid dehydratase (ilvD)	54.5%
AF1985	ketol-acid reductoisomerase (ilvC)	61.8%

Serine family

AF0813	phosphoglycerate dehydrogenase (serA)	48.8%
AF2138	phosphoserine phosphatase (serB)	50.7%
AF0273	sarcosine oxidase, subunit alpha (soxA)	31.1%
AF0274	sarcosine oxidase, subunit beta (soxB)	26.5%
AF0852	serine hydroxymethyltransferase (glyA)	56.1%

Histidine family

AF0590	ATP phosphoribosyltransferase (hisG)	31.8%
AF0212	histidinol dehydrogenase (hisD)	51.6%
AF2002	histidinol-phosphate aminotransferase (hisC-1)	39.8%
AF2024	histidinol-phosphate aminotransferase (hisC-2)	36.8%
AF0985	imidazoleglycerol-phosphate dehydrogenase/histidinol-phosphatase (hisB)	42.2%
AF0819	imidazoleglycerol-phosphate synthase, cyclase subunit (hisF)	67.0%
AF2265	imidazoleglycerol-phosphate synthase, subunit H (hisH)	44.4%
AF0509	imidazoleglycerol-phosphate synthase, subunit H, putative	43.2%
AF1960	phosphoribosyl-AMP cyclohydrolase/phosphoribosyl-ATP pyrophosphohydrolase (hisIE)	59.6%
AF0713	phosphoribosylformimino-5-aminoimidazole carboxamide ribotide isomerase (hisA-1)	37.5%
AF0988	phosphoribosylformimino-5-aminoimidazole carboxamide ribotide isomerase (hisA-2)	42.2%

BIOSYNTHESIS OF COFACTORS, PROSTHETIC GROUPS, AND CARRIERS

General

AF1855	2,3-dihydroxybenzoate-AMP ligase (entE)	27.2%
AF1070	coenzyme F390 synthetase (fstA-1)	30.3%
AF1671	coenzyme F390 synthetase (fstA-2)	31.9%
AF2013	coenzyme F390 synthetase (fstA-3)	30.4%
AF2151	isochorismatase (entB)	31.2%

Folic acid

AF1414	dihydropteroate synthase	40.8%

Heme and porphyrin

AF1648	bacteriochlorophyll synthase, 33 kDa subunit	27.9%
AF0464	bacteriochlorophyll synthase, 43 kDa subunit (chlP-1)	29.7%
AF1023	bacteriochlorophyll synthase, 43 kDa subunit (chlP-2)	31.2%
AF1637	bacteriochlorophyll synthase, 43 kDa subunit (chlP-3)	27.0%
AF0037	cobalamin (5'-phosphate) synthase (cobS-1)	33.9%
AF2323	cobalamin (5'-phosphate) synthase (cobS-2)	34.4%
AF0725	cobalamin biosynthesis precorrin methylase (cbiG)	30.7%
AF0727	cobalamin biosynthesis precorrin-2 methyltransferase (cbiL)	31.5%
AF0726	cobalamin biosynthesis precorrin-3 methylase (cbiF)	49.2%
AF0724	cobalamin biosynthesis precorrin-3 methylase (cbiH)	49.0%
AF0722	cobalamin biosynthesis precorrin-6Y methylase (cbiE)	32.4%
AF0732	cobalamin biosynthesis precorrin-8W decarboxylase (cbiT)	30.8%
AF1336	cobalamin biosynthesis protein (cbiB)	38.4%
AF0723	cobalamin biosynthesis protein (cbiD)	36.3%
AF0728	cobalamin biosynthesis protein (cbiM-1)	61.4%
AF1843	cobalamin biosynthesis protein (cbiM-2)	41.2%
AF0731	cobalt transport ATP-binding protein (cbiO-1)	47.2%
AF1841	cobalt transport ATP-binding protein (cbiO-2)	41.1%
AF0729	cobalt transport protein (cbiN)	56.0%
AF0730	cobalt transport protein (cbiQ-1)	32.6%
AF1842	cobalt transport protein (cbiQ-2)	30.3%
AF1338	cobyric acid synthase (cbiP)	44.5%
AF2229	cobyrinic acid a,c-diamide synthase (cbiA)	42.3%
AF1241	glutamate-1-semialdehyde aminotransferase (hemL)	54.3%
AF1975	glutamyl-tRNA reductase (hemA)	42.7%
AF1594	heme biosynthesis protein (nirH)	25.2%
AF1125	heme biosynthesis protein (nirJ-1)	38.7%
AF2009	heme biosynthesis protein (nirJ-2)	31.8%
AF1593	heme d1 biosynthesis protein (nirD)	29.4%
AF1311	oxygen-independent coproporphyrinogen III oxidase, putative	27.1%
AF1242	porphobilinogen deaminase (hemC)	46.3%
AF1974	porphobilinogen synthase (hemB)	60.4%
AF1784	protoporphyrinogen oxidase (hemK)	33.5%
AF0422	uroporphyrin-III C-methyltransferase (cysG-1)	41.7%
AF1243	uroporphyrin-III C-methyltransferase (cysG-2)	52.5%
AF0116	uroporphyrinogen III synthase (hemD)	27.4%

Menaquinone and ubiquinone

AF2176	4-hydroxybenzoate octaprenyltransferase (ubiA)	41.6%
AF0404	4-hydroxybenzoate octaprenyltransferase, putative	30.6%
AF2413	coenzyme PQQ synthesis protein (pqqE)	30.5%
AF1191	dihydroxynaphthoic acid synthase (menB)	54.6%
AF1561	octaprenyl-diphosphate synthase (lspB)	33.2%
AF0140	ubiquinone/menaquinone biosynthesis methyltransferase (ubiE)	31.0%

Molybdopterin

AF2006	molybdenum cofactor biosynthesis protein (moaA)	47.8%
AF0265	molybdenum cofactor biosynthesis protein (moaB)	44.4%
AF2150	molybdenum cofactor biosynthesis protein (moaC)	62.0%
AF0831	molybdenum cofactor biosynthesis protein (moaA-1)	50.8%
AF0930	molybdenum cofactor biosynthesis protein (moaA-2)	44.8%
AF0161	molybdenum cofactor biosynthesis protein (moaA-3)	30.5%
AF0531	molybdenum cofactor biosynthesis protein (moaB)	44.0%
AF1022	molybdenum-pterin-binding protein (mopB)	39.3%
AF1624	molybdopterin converting factor, subunit 1 (moaD)	36.6%
AF2179	molybdopterin converting factor, subunit 2 (moaE)	33.3%
AF2005	molybdopterin-guanine dinucleotide biosynthesis protein A (mobA)	33.2%
AF2253	molybdopterin-guanine dinucleotide biosynthesis protein B (mobB)	40.0%

Pantothenate

AF1645	pantothenate metabolism flavoprotein (dfp)	42.4%

Riboflavin

AF0484	GTP cyclohydrolase II (ribA-1)	44.5%
AF2107	GTP cyclohydrolase II (ribA-2)	47.1%
AF1416	riboflavin synthase (ribC)	53.3%
AF2128	riboflavin synthase, subunit beta (ribE)	75.9%
AF2007	riboflavin-specific deaminase (ribG)	43.7%

Thiamine

AF2075	hydroxyethylthiazole kinase (thiM)	33.6%
AF2208	hydroxymethylpyrimidine phosphate kinase (thiD)	35.5%
AF1695	thiamine biosynthesis protein (apbA)	36.9%
AF2412	thiamine biosynthesis protein (thiC)	60.2%
AF0553	thiamine biosynthesis protein (thiF)	38.1%
AF0088	thiamine biosynthesis protein, putative	28.2%
AF0702	thiamine biosynthetic enzyme (thi1)	50.0%
AF0733	thiamine monophosphate kinase (thiL)	30.4%
AF2074	thiamine phosphate pyrophosphorylase (thiE)	45.5%

Pyridine nucleotides

AF1000	NH(3)-dependent NAD+ synthetase (nadE)	52.0%
AF1839	nicotinate-nucleotide pyrophosphoryiase (nadC)	43.2%
AF1837	quinolinate synthetase (nadA), authentic frameshift	53.9%

CELL ENVELOPE

Membranes, lipoproteins, and porins

AF1420	membrane protein	51.9%
AF1354	membrane protein, putative	32.8%

Surface polysaccharides, lipopolysaccharides and antigens

AF0324	dTDP-glucose 4,6-dehydratase (rfbB)	50.0%
AF0043	first mannosyl transferase (wbaZ-1)	30.0%
AF0606	first mannosyl transferase (wbaZ-2)	29.0%
AF1728	galactosyltransferase	26.9%
AF0044	GDP-D-mannose dehydratase (gmd-1), authentic frameshift	40.7%
AF1142	glucose-1-phosphate cytidylyltransferase (rfbF)	38.6%
AF0242	glucose-1-phosphate thymidylyltransferase (graD-1)	27.7%
AF0325	glucose-1-phosphate thymidylyltransferase (graD-2)	45.2%
AF0321	glycosyl transferase	30.7%
AF0387	glycosyltransferase, putative	33.8%
AF0467	immunogenic protein (bcsp31-1)	34.7%
AF0635	immunogenic protein (bcsp31-2)	44.3%
AF0988	immunogenic protein (bcsp31-3)	28.3%
AF0602	LPS biosynthesis protein, putative	29.6%
AF0617	LPS biosynthesis protein, putative	29.0%
AF0607	LPS glycosyltransferase, putative	29.7%
AF0326	mannose-1-phosphate guanylyltransferase (rfbM), authentic frameshift	42.4%
AF1097	mannose-6-phosphate isomerase/mannose-1-phosphate guanylyl transferase (manC)	43.1%
AF0035	mannosephosphate isomerase, putative	31.3%
AF0045	mannosyltransferase A (mtfA)	38.7%
AF0311	O-antigen biosynthesis protein (rfbC), authentic frameshift	30.6%
AF0458	phosphomannomutase (pmm)	39.5%
AF0595	polysaccharide biosynthesis protein, putative	24.1%
AF0322	rhamnosyl transferase (rfbQ)	27.5%
AF0323	spore coat polysaccharide biosynthesis protein (spsK-2), authentic frameshift	36.3%
AF0620	succinoglycan biosynthesis protein (exoM)	24.8%
AF0361	UDP-glucose 4-epimerase (galE-1)	38.6%
AF2016	UDP-glucose 4-epimerase (galE-2)	30.0%
AF0302	UDP-glucose dehydrogenase (ugd-1)	43.8%
AF0596	UDP-glucose dehydrogenase (ugd-2)	44.1%

Surface structures

AF1054	flagellin (flaB1-1)	30.0%
AF1055	flagellin (flaB1-2)	31.1%
AF0275	surface layer protein B (slgB-1)	30.8%
AF1413	surface layer protein B (slgB-2)	29.9%

CELLULAR PROCESSES

General

AF1040	chemotaxis histidine kinase (cheA)	41.9%
AF1035	chemotaxis histidine kinase, putative	25.3%
AF1036	chemotaxis histidine kinase, putative	30.4%
AF1037	chemotaxis protein methyltransferase (cheR)	33.2%
AF1042	chemotaxis response regulator (cheY)	62.9%
AF1034	methyl-accepting chemotaxis protein (tlpC-1)	27.5%
AF1045	methyl-accepting chemotaxis protein (tlpC-2)	29.8%
AF1041	protein-glutamate methylesterase (cheB)	43.3%
AF1032	purine NTPase, putative	32.2%
AF1044	purine-binding chemotaxis protein (cheW)	40.4%

Cell division

AF0617	cell division control protein 21 (cdc21)	32.8%
AF1297	cell division control protein 48, AAA family (cdc48-1)	60.1%
AF2098	cell division control protein 48, AAA family (cdc48-2)	62.0%
AF0244	cell division control protein 6, putative	27.5%
AF1286	cell division control protein, AAA family, putative	49.3%
AF0696	cell division inhibitor (minD-1)	55.0%
AF1937	cell division inhibitor (minD-2)	32.8%
AF2051	cell division protein (ftsJ)	40.8%
AF0535	cell division protein (ftsZ-1)	60.4%
AF0570	cell division protein (ftsZ-2)	61.4%
AF0837	cell division protein pelota (pelA)	41.7%
AF1215	cell division protein, putative	32.8%
AF0238	centromere/microtubule-binding protein (cbf5)	58.8%
AF1558	chromosome segregation protein (smc1)	32.8%
AF1822	serine/threonine phosphatase (ppa)	31.9%

Chaperones

AF1298	small heat shock protein (hsp20-1)	52.3%
AF1971	small heat shock protein (hsp20-2)	38.1%
AF2238	thermosome, subunit alpha (thsA)	70.6%
AF1451	thermosome, subunit beta (thsB)	68.2%

Chromosome-associated protein

AF0337	archaeal histone A1 (hpyA1-1)	64.6%
AF1493	archaeal histone A1 (hpyA1-2)	69.7%

Detoxification

AF2173	2-nitropropane dioxygenase (ncd2)	39.7%
AF0270	alkyl hydroperoxide reductase	73.5%
AF1361	arsenate reductase (arsC)	30.5%
AF0650	N-ethylammeline chlorohydrolase (trzA-1)	45.9%
AF0997	N-ethylammeline chlorohydrolase (trzA-2)	44.5%
AF0254	NADH oxidase (noxA-1)	35.1%
AF0396	NADH oxidase (noxA-2)	35.5%
AF0400	NADH oxidase (noxA-3)	40.8%
AF0951	NADH oxidase (noxA-4)	36.7%
AF1858	NADH oxidase (noxA-5)	34.0%
AF0455	NADH oxidase (noxB-1)	43.3%
AF1262	NADH oxidase (noxB-2)	42.9%
AF0226	NADH oxidase (noxC)	38.4%
AF0515	NADH oxidase, putative	25.5%
AF2233	peroxidase / catalase (perA)	62.9%

Protein and peptide secretion

AF1902	protein translocase, subunit SEC61 alpha (secY)	50.0%
AF0536	protein translocase, subunit SEC61 gamma (secE)	25.0%
AF2062	signal recognition particle receptor (dpa)	54.8%
AF1258	signal recognition particle, subunit SRP19 (srp19)	36.6%
AF0622	signal recognition particle, subunit SRP54 (srp54)	51.2%
AF1791	signal sequence peptidase (sec11)	36.3%
AF1657	signal sequence peptidase (spc21)	47.0%
AF1655	signal sequence peptidase, putative	34.5%
AF0338	type II secretion system protein (gspE-1)	38.5%
AF0659	type II secretion system protein (gspE-2)	38.2%
AF0996	type II secretion system protein (gspE-3)	41.7%
AF1049	type II secretion system protein (gspE-4)	46.5%

CENTRAL INTERMEDIARY METABOLISM

Degradation of polysaccharides

AF1207	2-deoxy-D-gluconate 3-dehydrogenase (kduD)	45.3%
AF1796	endoglucanase (celM)	55.4%

Phosphorus compounds

AF0756	exopolyphosphatase (ppx1)	55.1%

Polyamine biosynthesis

AF0646	agmatinase (speB)	33.3%
AF2334	spermidine synthase (speE)	37.1%

Polysaccharides - (cytoplasmic)

AF0599	dolichol phosphate mannose synthase, putative	32.1%

Sulfur metabolism

AF0288	adenylylsulfate 3-phosphotransferase (cysC)	52.0%
AF1670	adenylylsulfate reductase, subunit A (aprA)	96.0%
AF1669	adenylylsulfate reductase, subunit B (aprB)	97.3%
AF1667	sulfate adenylyltransferase (sat)	28.4%
AF2228	sulfite reductase, desulfoviridin-type subunit gamma (dsvC)	41.3%
AF0423	sulfite reductase, subunit alpha (dsrA)	100.0%
AF0424	sulfite reductase, subunit beta (dsrB)	100.0%
AF0425	sulfite reductase, subunit gamma (dsrD)	97.4%

Other

AF1706	2-hydroxy-6-oxo-6-phenylhexa-2,4-dienoic acid hydrolase (pcbD)	29.4%
AF0675	2-hydroxy-6-oxohepta-2,4-dienoate hydrolase (todF)	26.3%
AF0091	2-hydroxyhepta-2,4-diene-1,7-dioate isomerase (hpcE-1)	44.5%
AF2225	2-hydroxyhepta-2,4-diene-1,7-dioate isomerase (hpcE-2)	66.0%
AF0333	4-hydroxyphenylacetate-3-hydroxylase (hpaA-1)	22.4%
AF0885	4-hydroxyphenylacetate-3-hydroxylase (hpaA-2)	26.0%
AF1027	4-hydroxyphenylacetate-3-hydroxylase (hpaA-3)	21.0%
AF0669	4-oxalocrotonate tautomerase, putative	31.9%
AF0808	glycolate oxidase subunit (glcD)	32.0%
AF2216	methylmalonyl-CoA decarboxylase, biotin carboxyl carrier subunit (mmdC)	36.2%
AF2217	methylmalonyl-CoA decarboxylase, subunit alpha (mmdA)	62.5%
AF1288	methylmalonyl-CoA mutase, subunit alpha (mutB), authentic frameshift	46.1%
AF2219	methylmalonyl-CoA mutase, subunit alpha, C-terminus (mcmA2)	48.7%
AF2215	methylmalonyl-CoA mutase, subunit alpha, N-terminus (mcmA1)	51.2%
AF2099	muconate cycloisomerase II (clcB)	24.9%
AF1425	phosphonopyruvate decarboxylase (bcpC-1)	35.0%
AF1751	phosphonopyruvate decarboxylase (bcpC-2)	48.6%

ENERGY METABOLISM

Amino acids and amines

AF1958	2-hydroxyglutaryl-CoA dehydratase, subunit alpha (hgdA)	30.5%

AF1957	2-hydroxyglutaryl-CoA dehydratase, subunit beta (hgdB)	24.4%
AF0130	acetylpolyamine aminohydrolase (aphA)	38.7%
AF2290	acetylpolyamine aminohydrolase, putative	33.3%
AF0991	glutaryl-CoA dehydrogenase (gcdH)	48.7%
AF1323	group II decarboxylase	28.0%
AF2004	group II decarboxylase	46.1%
AF2295	group II decarboxylase	30.5%
AF1665	ornithine cyclodeaminase (arcB)	36.3%
Anaerobic		
AF1145	4-hydroxybutyrate CoA transferase (cat2-1)	46.6%
AF1854	4-hydroxybutyrate CoA transferase (cat2-2)	47.6%
AF0866	glycerol kinase (glpK)	33.8%
AF1328	glycerol-3-phosphate dehydrogenase (glpA)	27.8%
AF0871	glycerol-3-phosphate dehydrogenase (NAD(P)+) (gpsA)	36.3%
AF0020	L-carnitine dehydratase (caiB-1)	33.3%
AF0990	L-carnitine dehydratase (caiB-2)	31.2%
ATP-proton motive force interconversion		
AF1158	ATP synthase, subunit E, putative	47.1%
AF1166	H+-transporting ATP synthase, subunit A (atpA)	67.0%
AF1167	H+-transporting ATP synthase, subunit B (atpB)	72.6%
AF1164	H+-transporting ATP synthase, subunit C (atpC)	37.5%
AF1168	H+-transporting ATP synthase, subunit D (atpD)	47.1%
AF1163	H+-transporting ATP synthase, subunit E (atpE)	36.3%
AF1165	H+-transporting ATP synthase, subunit F (atpF)	46.0%
AF1159	H+-transporting ATP synthase, subunit I (atpI)	30.1%
AF1160	H+-transporting ATP synthase, subunit K (atpK-1)	46.3%
AF1162	H+-transporting ATP synthase, subunit K (atpK-2)	46.3%
Electron transport		
AF2036	cytochrome C oxidase folding protein (coxD)	33.3%
AF0144	cytochrome C oxidase, subunit II (cbaB)	34.2%
AF0142	cytochrome C oxidase, subunit II, putative	38.0%
AF0190	cytochrome C oxidase, subunit II, putative	31.7%
AF1057	cytochrome C-type biogenesis protein (ccdA)	30.7%
AF2192	cytochrome C-type biogenesis protein (nrfE)	36.1%
AF2296	cytochrome oxidase, subunit I (cydA-1)	22.9%
AF2297	cytochrome oxidase, subunit I (cydA-2)	31.5%
AF2046	cytochrome oxidase, subunit I, putative	25.1%
AF0528	cytochrome-c3 hydrogenase, subunit gamma	39.3%
AF0833	desulfoferrodoxin (dfx)	63.0%
AF0344	desulfoferrodoxin, putative	47.3%
AF0287	electron transfer flavoprotein, subunit alpha (etfA)	39.7%
AF0286	electron transfer flavoprotein, subunit beta (etfB)	38.8%
AF1380	F420-nonreducing hydrogenase (vhtA)	34.8%
AF1371	F420-nonreducing hydrogenase (vhtD-1)	30.9%
AF1378	F420-nonreducing hydrogenase (vhtD-2)	33.1%
AF1381	F420-nonreducing hydrogenase (vhtG)	46.1%
AF1824	F420H2:quinone oxidoreductase, 11.2 kDa subunit, putative	24.1%
AF1823	F420H2:quinone oxidoreductase, 16.5 kDa subunit, putative	25.7%
AF1832	F420H2:quinone oxidoreductase, 32 kDa subunit (nuoI)	95.6%
AF1833	F420H2:quinone oxidoreductase, 39 kDa subunit, putative	33.6%
AF1829	F420H2:quinone oxidoreductase, 39.7 kDa subunit, putative	43.8%
AF1831	F420H2:quinone oxidoreductase, 41.2 kDa subunit, putative	34.8%
AF1827	F420H2:quinone oxidoreductase, 43.2 kDa subunit, putative	26.9%
AF1830	F420H2:quinone oxidoreductase, 45 kDa subunit (nuoD)	80.0%
AF1825	F420H2:quinone oxidoreductase, 53.9 kDa subunit (nuoM)	32.1%
AF1826	F420H2:quinone oxidoreductase, 72.4 kDa subunit (nuoL)	33.2%
AF0158	ferredoxin (fdx-1)	45.3%
AF0166	ferredoxin (fdx-2)	49.2%
AF0355	ferredoxin (fdx-3)	53.2%
AF0427	ferredoxin (fdx-4)	56.1%
AF0923	ferredoxin (fdx-5)	56.9%
AF1010	ferredoxin (fdx-6)	44.4%
AF1239	ferredoxin (fdx-7)	29.0%
AF2142	ferredoxin (fdx-8)	38.0%
AF0164	ferredoxin-nitrite reductase (nirA)	29.7%
AF2332	flavodoxin, putative	30.3%
AF0167	flavoprotein (fprA-1)	33.2%
AF1520	flavoprotein (fprA-2)	47.2%
AF0557	flavoprotein reductase	26.2%
AF1463	fumarate reductase, flavoprotein subunit (frdA)	27.0%
AF1536	glutaredoxin (grx-1)	34.3%
AF2145	glutaredoxin (grx-2)	38.8%
AF0663	heterodisulfide reductase, subunit A (hdrA-1)	42.2%
AF1377	heterodisulfide reductase, subunit A (hdrA-2)	46.8%
AF0662	heterodisulfide reductase, subunit A/ methylviologen reducing hydrogenase, subunit delta	34.2%
AF1238	heterodisulfide reductase, subunit A/methylviologen reducing hydrogenase, subunit delta	53.7%
AF1375	heterodisulfide reductase, subunit B (hdrB)	36.0%
AF0271	heterodisulfide reductase, subunit B, putative	35.3%
AF1376	heterodisulfide reductase, subunit C (hdrC)	33.3%
AF0502	heterodisulfide reductase, subunit D, putative	33.8%
AF0809	heterodisulfide reductase, subunit D, putative	100.0%
AF0661	heterodisulfide reductase, subunit E, putative	23.8%
AF0755	heterodisulfide reductase, subunits E and D, putative	31.8%
AF0506	iron-sulfur binding reductase	38.5%
AF1773	iron-sulfur binding reductase	33.3%
AF1998	iron-sulfur binding reductase	29.6%
AF0627	iron-sulfur cluster binding protein	45.5%
AF0688	iron-sulfur cluster binding protein	44.8%
AF1153	iron-sulfur cluster binding protein	27.9%
AF1185	iron-sulfur cluster binding protein	36.7%
AF1263	iron-sulfur cluster binding protein	42.1%
AF2380	iron-sulfur cluster binding protein	35.3%
AF2381	iron-sulfur cluster binding protein	34.4%
AF2409	iron-sulfur cluster binding protein	28.2%
AF0076	iron-sulfur cluster binding protein	32.7%
AF1461	iron-sulfur cluster binding protein, putative	51.0%
AF1436	iron-sulfur flavoprotein (isf-1)	35.7%
AF1519	iron-sulfur flavoprotein (isf-2)	56.6%
AF1896	iron-sulfur flavoprotein (isf-3)	37.1%
AF1372	methylviologen-reducing hydrogenase, subunit alpha (vhuA)	39.4%
AF1374	methylviologen-reducing hydrogenase, subunit delta (vhuD)	41.7%
AF1373	methylviologen-reducing hydrogenase, subunit gamma (vhuG)	38.6%
AF0157	molybdopterin oxidoreductase, iron-sulfur binding subunit	38.6%
AF0174	molybdopterin oxidoreductase, membrane subunit	26.0%
AF0175	molybdopterin oxidoreductase, iron-sulfur binding subunit	42.0%
AF0176	molybdopterin oxidoreductase, molybdopterin binding subunit	32.6%
AF0499	molybdopterin oxidoreductase, iron-sulfur binding subunit	41.5%
AF0500	molybdopterin oxidoreductase, membrane subunit	27.9%
AF1202	molybdopterin oxidoreductase, iron-sulfur binding subunit	35.5%
AF1203	molybdopterin oxidoreductase, molybdopterin binding subunit	30.1%
AF2384	molybdopterin oxidoreductase, molybdopterin binding subunit	34.6%
AF2385	molybdopterin oxidoreductase, iron-sulfur binding subunit	46.9%
AF2386	molybdopterin oxidoreductase, membrane subunit	30.3%
AF0159	molybdopterin oxidoreductase, molybdopterin binding subunit, putative	30.9%
AF2267	NAD(P)H-flavin oxidoreductase	31.4%
AF0131	NAD(P)H-flavin oxidoreductase, putative	28.2%
AF2352	NADH dehydrogenase, subunit I, putative	28.9%
AF1828	NADH dehydrogenase, subunit 3	24.3%
AF0248	NADH-dependent flavin oxidoreductase	36.7%
AF0342	nigerythrin, putative	33.3%
AF0546	nitrate reductase, gamma subunit (narI)	30.1%
AF0501	nitrate reductase, gamma subunit, putative	29.3%
AF1126	P450 cytochrome, putative	30.5%
AF0463	polyferredoxin (mvhB), authentic frameshift	32.2%
AF1379	quinone-reactive Ni/Fe-hydrogenase B-type cytochrome subunit (hydC)	29.0%
AF0173	reductase, assembly protein	30.0%
AF0547	reductase, iron-sulfur binding subunit	28.3%
AF0867	reductase, putative	33.3%
AF0880	rubredoxin (rd-1)	69.2%
AF1349	rubredoxin (rd-2)	67.9%
AF0832	rubrerythrin (rr1)	45.7%
AF0831	rubrerythrin (rr2)	63.7%
AF1640	rubrerythrin (rr3)	37.8%
AF2312	rubrerythrin (rr4)	41.4%
AF0711	thioredoxin (trx-1)	28.4%
AF0769	thioredoxin (trx-2)	38.5%
AF1284	thioredoxin (trx-3)	52.9%
AF2144	thioredoxin (trx-4)	48.9%
AF1339	ubiquinol-cytochrome C reductase complex, subunit VI requiring protein	60.9%
Fermentation		
AF1779	2-hydroxyacid dehydrogenase, putative	37.6%
AF0469	2-ketoglutarate ferredoxin oxidoreductase, subunit alpha (korA)	52.3%
AF0468	2-ketoglutarate ferredoxin oxidoreductase, subunit beta (korB)	51.2%
AF0470	2-ketoglutarate ferredoxin oxidoreductase, subunit delta (korD)	47.2%
AF0471	2-ketoglutarate ferredoxin oxidoreductase, subunit gamma (korG)	40.0%
AF2053	2-ketoisovalerate ferredoxin oxidoreductase, subunit alpha (vorA)	41.2%
AF2052	2-ketoisovalerate ferredoxin oxidoreductase, subunit beta (vorB)	42.7%
AF2054	2-ketoisovalerate ferredoxin oxidoreductase, subunit delta (vorD)	51.5%
AF2055	2-ketoisovalerate ferredoxin oxidoreductase, subunit gamma (vorG)	45.2%
AF0749	2-oxoacid ferredoxin oxidoreductase, subunit alpha (orA)	33.7%
AF0750	2-oxoacid ferredoxin oxidoreductase, subunit beta (orB)	49.2%
AF1286	acetoin utilization protein, putative	35.1%
AF0197	acetyl-CoA synthetase (acs-1)	27.1%
AF0366	acetyl-CoA synthetase (acs-2)	47.3%
AF0677	acetyl-CoA synthetase (acs-3)	40.9%
AF0975	acetyl-CoA synthetase (acs-4)	42.3%
AF0976	acetyl-CoA synthetase (acs-6)	36.2%
AF1287	acetyl-CoA synthetase (acs-6)	34.3%
AF0024	alcohol dehydrogenase, iron-containing	36.2%
AF0339	alcohol dehydrogenase, iron-containing	37.4%
AF2019	alcohol dehydrogenase, iron-containing	35.7%
AF2389-C	acetyl-CoA synthetase, putative	64.8%
AF2389-N	acetyl-CoA synthetase, putative	59.3%
AF2101	alcohol dehydrogenase, zinc-dependent	34.8%
AF0023	aldehyde ferredoxin oxidoreductase (aor-1)	41.1%
AF0077	aldehyde ferredoxin oxidoreductase (aor-2)	32.6%
AF0340	aldehyde ferredoxin oxidoreductase (aor-3)	38.4%
AF2281	aldehyde ferredoxin oxidoreductase (aor-4)	53.0%
AF0006	corrinoid methyltransferase protein (mtaC-1)	30.7%
AF0011	corrinoid methyltransferase protein (mtaC-2)	29.5%
AF0394	D-lactate dehydrogenase, cytochrome-type (dld)	31.9%
AF0560	formate dehydrogenase (fdhD1), authentic frameshift	32.9%
AF1199	glutaconate CoA-transferase, subunit A (gctA)	31.9%
AF1198	glutaconate CoA-transferase, subunit B (gctB), authentic frameshift	37.0%
AF1489	indolepyruvate ferredoxin oxidoreductase, subunit alpha (iorA)	48.1%
AF2030	indolepyruvate ferredoxin oxidoreductase, subunit beta (iorB)	41.1%
AF0807	L-lactate dehydrogenase, cytochrome-type (lldD)	39.4%
AF0855	L-malate dehydrogenase, NAD+-dependent (mdhA)	40.1%
AF2085	oxaloacetate decarboxylase, biotin carboxyl carrier subunit, putative	38.7%
AF2084	oxaloacetate decarboxylase, sodium ion pump subunit (oadB)	59.8%
AF1252	oxaloacetate decarboxylase, subunit alpha (oadA)	63.3%
AF1701	pyruvate ferredoxin oxidoreductase, subunit alpha (porA)	50.3%
AF1702	pyruvate ferredoxin oxidoreductase, subunit beta (porB)	50.7%
AF1700	pyruvate ferredoxin oxidoreductase, subunit delta (porD)	53.1%
AF1699	pyruvate ferredoxin oxidoreductase, subunit gamma (porG)	50.8%
Gluconeogenesis		
AF0710	phosphoenolpyruvate synthase (ppsA)	61.4%
Glycolysis		
AF1146	3-phosphoglycerate kinase (pgk)	48.8%
AF1132	enolase (eno)	53.9%
AF1732	glyceraldehyde 3-phosphate dehydrogenase (gap)	56.6%
AF1304	triosephosphate isomerase (tpiA)	56.4%
Pentose phosphate pathway		
AF0943	ribose 5-phosphate isomerase (rpi)	48.9%
Sugars		
AF0356	carbohydrate kinase, pfkB family	31.3%
AF0401	carbohydrate kinase, pfkB family	34.1%
AF1324	carbohydrate kinase, FGGY family	27.1%
AF1752	carbohydrate kinase, FGGY family	29.3%
AF0861	D-arabino 3-hexulose 6-phosphate formaldehyde lyase (hps-1)	30.6%
AF1305	D-arabino 3-hexulose 6-phosphate formaldehyde lyase (hps-2)	44.2%
AF0480	fuculose-1-phosphate aldolase (fucA)	31.8%
TCA cycle		
AF1963	aconitase (acn)	57.1%
AF1340	citrate synthase (citZ)	50.3%
AF1098	fumarase (fum-1)	49.1%
AF1099	fumarase (fum-2)	53.4%
AF0647	isocitrate dehydrogenase, NADP (icd)	57.2%
AF1727	malate oxidoreductase (mae)	52.3%
AF0681	succinate dehydrogenase, flavoprotein subunit A (sdhA)	48.2%
AF0682	succinate dehydrogenase, iron-sulfur subunit B (sdhB)	51.3%
AF0683	succinate dehydrogenase, subunit C (sdhC)	38.6%
AF0684	succinate dehydrogenase, subunit D (sdhD)	25.9%
AF1539	succinyl-CoA synthetase, alpha subunit (sucD-1)	56.9%
AF2185	succinyl-CoA synthetase, alpha subunit (sucD-2)	63.5%
AF1540	succinyl-CoA synthetase, beta subunit (sucC-1)	51.3%
AF2186	succinyl-CoA synthetase, beta subunit (sucC-2)	49.8%

FATTY ACID AND PHOSPHOLIPID METABOLISM

General		
AF1736	3-hydroxy-3-methylglutaryl-coenzyme A reductase (mvaA)	57.1%
AF0017	3-hydroxyacyl-CoA dehydrogenase (hbd-1)	41.1%
AF0285	3-hydroxyacyl-CoA dehydrogenase (hbd-2)	55.8%
AF0434	3-hydroxyacyl-CoA dehydrogenase (hbd-3)	40.7%
AF1025	3-hydroxyacyl-CoA dehydrogenase (hbd-4)	46.6%
AF1122	3-hydroxyacyl-CoA dehydrogenase (hbd-5)	45.2%
AF1177	3-hydroxyacyl-CoA dehydrogenase (hbd-6)	36.8%
AF1190	3-hydroxyacyl-CoA dehydrogenase (hbd-7)	46.5%
AF1206	3-hydroxyacyl-CoA dehydrogenase (hbd-8)	36.3%
AF2017	3-hydroxyacyl-CoA dehydrogenase (hbd-9)	35.4%
AF2273	3-hydroxyacyl-CoA dehydrogenase (hbd-10)	39.4%
AF0018	3-ketoacyl-CoA thiolase (acaB-1)	41.0%
AF0034	3-ketoacyl-CoA thiolase (acaB-2)	38.3%
AF0133	3-ketoacyl-CoA thiolase (acaB-3)	32.3%
AF0134	3-ketoacyl-CoA thiolase (acaB-4)	32.5%
AF0201	3-ketoacyl-CoA thiolase (acaB-5)	26.9%
AF0202	3-ketoacyl-CoA thiolase (acaB-6)	33.5%
AF0283	3-ketoacyl-CoA thiolase (acaB-7)	42.0%
AF0438	3-ketoacyl-CoA thiolase (acaB-8)	42.4%
AF0967	3-ketoacyl-CoA thiolase (acaB-9)	33.7%
AF0968	3-ketoacyl-CoA thiolase (acaB-10)	28.0%
AF1291	3-ketoacyl-CoA thiolase (acaB-11)	40.1%
AF2416	3-ketoacyl-CoA thiolase (acaB-12)	49.9%
AF1028	3-ketoacyl-CoA thiolase (fadA-1)	38.8%
AF1197	3-ketoacyl-CoA thiolase (fadA-2)	47.2%
AF2243	3-ketoacyl-CoA thiolase (fadA-3)	40.3%
AF0033	acyl carrier protein synthase (acaA-1)	28.6%
AF2415	acyl carrier protein synthase (acaA-2)	58.7%
AF0199	acyl-CoA dehydrogenase (acd-1)	35.9%
AF0436	acyl-CoA dehydrogenase (acd-2)	44.1%
AF0498	acyl-coA dehydrogenase (acd-3)	22.9%
AF0671	acyl-CoA dehydrogenase (acd-4)	37.9%
AF0845	acyl-CoA dehydrogenase (acd-5)	44.6%
AF0964	acyl-CoA dehydrogenase (acd-6)	35.8%
AF1026	acyl-CoA dehydrogenase (acd-7)	42.6%
AF1141	acyl-CoA dehydrogenase (acd-8)	43.2%
AF1293	acyl-CoA dehydrogenase (acd-9)	45.8%
AF2057	acyl-CoA dehydrogenase (acd-10)	44.8%
AF2244	acyl-CoA dehydrogenase (acd-11)	42.6%
AF2275	acyl-CoA dehydrogenase (acd-12)	38.9%
AF1175	acyl-CoA dehydrogenase, short chain-specific (acdS)	30.1%
AF0818	acylphosphatase (acyP)	36.8%
AF0868	alkyldihydroxyacetonephosphate synthase	33.6%
AF2286	bifunctional short chain isoprenyl diphosphate synthase (idsA)	42.7%
AF0220	biotin carboxylase (acc)	59.1%
AF0865	carboxylesterase (est-1)	27.1%
AF1537	carboxylesterase (est-2)	29.0%
AF2336	carboxylesterase (est-3)	30.4%
AF1716	carboxylesterase (estA)	40.4%
AF1744	CDP-diacylglycerol--glycerol-3-phosphate 3-phosphatidyltransferase (pgsA-2)	26.7%
AF1143	CDP-diacylglycerol--glycerol-3-phosphate 3-phosphatidyltransferase (pgsA-1)	27.0%
AF2044	CDP-diacylglycerol--serine O-phosphatidyltransferase (pssA)	36.6%
AF0435	enoyl-CoA hydratase (fad-1)	47.6%
AF0685	enoyl-CoA hydratase (fad-2)	39.9%
AF0963	enoyl-CoA hydratase (fad-3)	48.6%
AF1641	enoyl-CoA hydratase (fad-4)	41.7%
AF2429	enoyl-CoA hydratase (fad-5)	34.7%
AF1763	lipase, putative	33.5%
AF0089	long-chain-fatty-acid--CoA ligase (fadD-1)	31.9%
AF0200	long-chain-fatty-acid--CoA ligase (fadD-2)	34.8%
AF0687	long-chain-fatty-acid--CoA ligase (fadD-3)	31.1%
AF0840	long-chain-fatty-acid--CoA ligase (fadD-4)	38.1%
AF1029	long-chain-fatty-acid--CoA ligase (fadD-5)	37.8%
AF1510	long-chain-fatty-acid--CoA ligase (fadD-6)	36.0%
AF1772	long-chain-fatty-acid--CoA ligase (fadD-7)	38.7%
AF1932	long-chain-fatty-acid--CoA ligase (fadD-8)	31.0%
AF2368	long-chain-fatty-acid--CoA ligase (fadD-9)	38.7%
AF1753	lysophospholipase	33.5%
AF0196	medium-chain acyl-CoA ligase (alkK-1)	34.6%
AF0262	medium-chain acyl-CoA ligase (alkK-2)	38.6%
AF0672	medium-chain acyl-CoA ligase (alkK-3)	31.0%
AF1261	medium-chain acyl-CoA ligase (alkK-4)	42.7%
AF2033	medium-chain acyl-CoA ligase (alkK-5)	33.5%
AF2289	mevalonate kinase (mvk)	40.6%
AF1794	myo-inositol-1-phosphate synthase (ino1)	32.2%
AF2045	phosphatidylserine decarboxylase (psd2)	42.5%
AF1674	sn-glycerol-1-phosphate dehydrogenase (gldA)	44.0%

AUTOTROPHIC METABOLISM

General		
AF1100	acetyl-CoA decarbonylase/synthase, subunit alpha (cdhA-1)	50.4%
AF2397	acetyl-CoA decarbonylase/synthase, subunit alpha (cdhA-2)	54.0%
AF0379	acetyl-CoA decarbonylase/synthase, subunit beta (cdhC)	62.7%
AF0377	acetyl-CoA decarbonylase/synthase, subunit delta (cdhD)	57.4%
AF1101	acetyl-CoA decarbonylase/synthase, subunit epsilon (cdhB-1)	40.0%
AF2398	acetyl-CoA decarbonylase/synthase, subunit epsilon (cdhB-2)	38.9%
AF0376	acetyl-CoA decarbonylase/synthase, subunit gamma (cdhE)	55.4%
AF1849	carbon monoxide dehydrogenase, catalytic subunit (cooS)	39.9%
AF0950	carbon monoxide dehydrogenase, iron sulfur subunit (cooF)	38.9%
AF1535	ferredoxin-thioredoxin reductase, catalytic subunit (ftrB)	38.6%
AF2073	formylmethanofuran:tetrahydromethanopterin formyltransferase (ftr-1)	46.0%
AF2207	formylmethanofuran:tetrahydromethanopterin formyltransferase (ftr-2)	68.4%

AF1935	N5,N10-methenyltetrahydromethanopterin cyclohydrolase (mch)	97.3%
AF0714	N5,N10-methylenetetrahydromethanopterin dehydrogenase (mtd)	61.8%
AF1066	N5,N10-methylenetetrahydromethanopterin reductase (mer-1)	59.1%
AF1196	N5,N10-methylenetetrahydromethanopterin reductase (mer-2)	37.4%
AF0009	N5-methyltetrahydromethanopterin:coenzyme M methyltransferase (mtr)	42.1%
AF1587	ribulose bisphosphate carboxylase, large subunit (rbcL-1)	40.6%
AF1638	ribulose bisphosphate carboxylase, large subunit (rbcL-2)	44.9%
AF1930	tungsten formylmethanofuran dehydrogenase, subunit A (fwdA)	48.9%
AF1650	tungsten formylmethanofuran dehydrogenase, subunit B (fwdB-1)	37.0%
AF1929	tungsten formylmethanofuran dehydrogenase, subunit B (fwdB-2)	49.4%
AF1931	tungsten formylmethanofuran dehydrogenase, subunit C (fwdC)	44.1%
AF1651	tungsten formylmethanofuran dehydrogenase, subunit D (fwdD-1)	32.6%
AF1928	tungsten formylmethanofuran dehydrogenase, subunit D (fwdD-2)	52.6%
AF0177	tungsten formylmethanofuran dehydrogenase, subunit E (fwdE)	29.7%
AF1644	tungsten formylmethanofuran dehydrogenase, subunit F (fwdF)	38.2%
AF1649	tungsten formylmethanofuran dehydrogenase, subunit G (fwdG)	45.6%

PURINES, PYRIMIDINES, NUCLEOSIDES, AND NUCLEOTIDES

2'-Deoxyribonucleotide metabolism

AF1108	deoxycytidine triphosphate deaminase, putative	38.1%
AF1664	ribonucleotide reductase (nrd)	59.7%
AF1554	thioredoxin reductase (trxB)	45.2%
AF2047	thymidylate synthase, putative	33.1%

Nucleotide and nucleaside interconversions

AF0676	5'-nucleotidase (nt5)	30.9%
AF0676	adenylate kinase (adk)	56.1%
AF1900	cytidylate kinase (cmk)	48.6%
AF0767	nucleoside diphosphate kinase (ndk)	56.4%
AF0061	thymidylate kinase (tmk)	34.9%
AF1308	thymidylate kinase, putative	26.3%
AF2042	uridylate kinase (pyrH)	53.6%

Purine ribonucleotide biosynthesis

AF2242	adenylosuccinate lyase (purB)	52.3%
AF0841	adenylosuccinate synthetase (purA)	70.8%
AF0673	amidophosphoribosyltransferase (purF)	56.8%
AF0253	GMP synthase (guaA-1)	59.8%
AF1320	GMP synthase (guaA-2)	49.4%
AF1811	inosine monophosphate cyclohydrolase	38.3%
AF0847	inosine monophosphate dehydrogenase (guaB-1)	41.6%
AF2118	inosine monophosphate dehydrogenase (guaB-2)	31.9%
AF1259	inosine monophosphate dehydrogenase, putative	51.8%
AF1157	phosphoribosylamine--glycine ligase (purD)	40.9%
AF1271	phosphoribosylaminoimidazole carboxylase (purE)	42.8%
AF1272	phosphoribosylaminoimidazolesuccinocarboxamide synthase (purC)	34.7%
AF1693	phosphoribosylformylglycinamidine cyclo-ligase (purM)	53.8%
AF1260	phosphoribosylformylglycinamidine synthase I (purQ)	40.9%
AF1940	phosphoribosylformylglycinamidine synthase II (purL)	41.5%
AF0589	ribose-phosphate pyrophosphokinase (prsA-1)	35.0%
AF1419	ribose-phosphate pyrophosphokinase (prsA-2)	41.1%

Pyrimidine ribonucleotide biosynthesis

AF0106	aspartate carbamoyltransferase, catalytic subunit (pyrB)	60.7%
AF0107	aspartate carbamoyltransferase, regulatory subunit (pyrI)	48.2%
AF1274	carbamoyl-phosphate synthase, large subunit (carB)	65.1%
AF1273	carbamoyl-phosphate synthase, small subunit (carA)	55.2%
AF0252	CTP synthase (pyrG)	58.3%
AF2250	dihydroorotase (pyrC)	37.2%
AF0745	dihydroorotase dehydrogenase (pyrD)	44.8%
AF1741	orotate phosphoribosyl transferase (pyrE)	49.0%
AF0306	orotate phosphoribosyl transferase, putative	39.0%

Salvage of nucleosides and nucleotides

AF0240	adenine deaminase (adeC)	39.5%
AF1764	dCMP deaminase, putative	39.0%
AF1788	methylthioadenosine phosphorylase (mtaP)	40.0%
AF1341	thymidine phosphorylase (deoA-1)	46.7%
AF1342	thymidine phosphorylase (deoA-2)	40.7%
AF0239	xanthine-guanine phosphoribosyltransferase (gptA-1)	25.7%
AF1789	xanthine-guanine phosphoribosyltransferase (gptA-2)	28.2%

REGULATORY FUNCTIONS

AF1959	(R)-hydroxyglutaryl-CoA dehydratase activator (hgdC)	51.2%
AF0168	arsenical resistance operon repressor, putative	36.7%
AF2204	arylsulfatase regulatory protein, putative	29.9%
AF0074	biotin operon repressor/biotin--[acetyl CoA carboxylase] ligase (birA)	36.6%
AF1724	dinitrogenase reductase activating glycohydrolase (draG)	37.9%
AF2232	ferric uptake regulation protein (fur)	25.8%
AF1785	iron-dependent repressor	42.0%
AF2395	iron-dependent repressor	40.0%
AF0245	iron-dependent repressor (desR)	28.2%
AF1984	iron-dependent repressor (troR)	28.3%
AF2430	lacZ expression regulatory protein (icc)	29.6%
AF1622	leucine responsive regulatory protein (lrp)	29.1%
AF0673	mercuric resistance operon regulatory protein (merR)	37.6%
AF2425	methanol dehydrogenase regulatory protein (moxR)	48.3%
AF1475	mitochondrial benzodiazepine receptor/sensory transduction protein	38.4%
AF0198	monoamine oxidase regulatory protein, putative	41.7%
AF1933	monoamine oxidase regulatory protein, putative	38.9%
AF0978	nitrogen regulatory protein P-II (glnB-1)	61.7%
AF1747	nitrogen regulatory protein P-II (glnB-2)	58.0%
AF1750	nitrogen regulatory protein P-II (glnB-3)	60.7%
AF0331	pheromone shutdown protein (traB)	40.5%
AF1797	phosphate regulatory protein, putative	30.7%
AF0521	protease synthase and sporulation regulator Pai1, putative	52.4%
AF1627	repressor protein	59.1%
AF1793	repressor protein	54.5%
AF0449	response regulator	38.1%
AF1063	response regulator	36.3%
AF1256	response regulator	42.5%
AF1384	response regulator	44.7%
AF1473	response regulator	32.5%
AF1898	response regulator	48.7%
AF2249	response regulator	44.8%
AF2419	response regulator	37.9%
AF0004	RNase L inhibitor	54.5%
AF0021	signal-transducing histidine kinase	26.1%
AF0208	signal-transducing histidine kinase	27.9%
AF0450	signal-transducing histidine kinase	32.4%
AF0770	signal-transducing histidine kinase	26.9%
AF0893	signal-transducing histidine kinase	28.7%
AF1184	signal-transducing histidine kinase	29.8%
AF1452	signal-transducing histidine kinase	28.5%
AF1467	signal-transducing histidine kinase	37.4%
AF1472	signal-transducing histidine kinase	30.4%
AF1483	signal-transducing histidine kinase	27.7%
AF1515	signal-transducing histidine kinase	32.0%
AF1639	signal-transducing histidine kinase	29.9%
AF1721	signal-transducing histidine kinase	34.5%
AF2109	signal-transducing histidine kinase	31.6%
AF0881	signal-transducing histidine kinase, authentic frameshift	26.5%
AF0277	signal-transducing histidine kinase, putative	29.8%
AF0410	signal-transducing histidine kinase, putative	27.1%
AF0448	signal-transducing histidine kinase, putative	26.1%
AF1820	signal-transducing histidine kinase, putative	26.2%
AF2032	signal-transducing histidine kinase, putative	22.5%
AF2420	signal-transducing histidine kinase, putative	28.4%
AF0442	succinoglycan biosynthesis regulator (exsB)	37.2%
AF1516	sugar fermentation stimulation protein (sfsA)	31.0%
AF1270	transcriptional regulatory protein, ArsR family	35.4%
AF1544	transcriptional regulatory protein, ArsR family	32.3%
AF1853	transcriptional regulatory protein, ArsR family	34.9%
AF2136	transcriptional regulatory protein, ArsR family	39.8%
AF0439	transcriptional regulatory protein, AsnC family	29.8%
AF0474	transcriptional regulatory protein, AsnC family	51.0%
AF0584	transcriptional regulatory protein, AsnC family	35.3%
AF1121	transcriptional regulatory protein, AsnC family	35.8%
AF1148	transcriptional regulatory protein, AsnC family	32.6%
AF1404	transcriptional regulatory protein, AsnC family	45.1%
AF1448	transcriptional regulatory protein, AsnC family	30.6%
AF1723	transcriptional regulatory protein, AsnC family	46.4%
AF1743	transcriptional regulatory protein, AsnC family	34.9%
AF2127	transcriptional regulatory protein, LysR family	30.8%
AF0114	transcriptional regulatory protein, putative	35.6%
AF1968	transcriptional regulatory protein, Rok family	32.9%
AF0112	transcriptional regulatory protein, Sir2 family	38.9%
AF1676	transcriptional regulatory protein, Sir2 family	40.8%
AF1817	transcriptional regulatory protein, TetR family	24.5%
AF0363	transcriptional repressor (cinR)	27.5%

REPLICATION

DNA replication, restriction, modification, recombination, and repair

AF2117	3-methyladenine DNA glycosylase (alkA)	30.0%
AF2060	activator 1, replication factor C, 35 KDa subunit	66.3%
AF1195	activator 1, replication factor C, 53 KDa subunit	43.7%
AF0465	DNA gyrase, subunit A (gyrA)	48.4%
AF0530	DNA gyrase, subunit B (gyrB)	58.4%
AF1388	DNA helicase, putative	46.8%
AF1960	DNA helicase, putative	32.7%
AF0623	DNA ligase (lig)	44.4%
AF1725	DNA ligase, putative	32.7%
AF0497	DNA polymerase B1 (polB)	45.1%
AF0693	DNA polymerase B2 (boxA), authentic frameshift	32.3%
AF0972	DNA polymerase III, subunit epsilon (dnaQ)	31.9%
AF2277	DNA polymerase, bacteriophage-type	36.9%
AF0742	DNA primase, putative	26.8%
AF0264	DNA repair protein RAD2 (rad2)	44.4%
AF0358	DNA repair protein RAD25	32.5%
AF1031	DNA repair protein RAD32 (rad32)	37.6%
AF0993	DNA repair protein RAD51 (radA)	59.3%
AF2096	DNA repair protein REC	40.0%
AF2418	DNA repair protein, putative	28.9%
AF1806	DNA topoisomerase I (topA)	36.2%
AF0940	DNA topoisomerase VI, subunit A (top6A)	39.8%
AF0652	DNA topoisomerase VI, subunit B (top6B)	43.9%
AF1892	endonuclease III (nth)	44.3%
AF0580	exodeoxyribonuclease III (xthA)	41.3%
AF2314	methylated-DNA-protein-cysteine methyltransferase (ogt)	55.3%
AF1409	modification methylase, type III R/M system	31.4%
AF1234	mutator protein MutT (mutT)	63.6%
AF2200	mutator protein MutT, putative	42.0%
AF0335	proliferating-cell nuclear antigen (pol30)	33.7%
AF0694	replication control protein A, putative	30.2%
AF1024	reverse gyrase (top-RG)	40.7%
AF0621	ribonuclease HII (rnhB)	39.3%
AF1715	type I restriction-modification enzyme, M subunit, authentic frameshift	63.0%
AF1708	type I restriction-modification enzyme, R subunit	38.2%
AF1710	type I restriction-modification enzyme, S subunit	33.0%

TRANSCRIPTION

DNA-dependent RNA polymerase

AF1888	DNA-directed RNA polymerase, subunit A' (rpoA1)	63.6%
AF1889	DNA-directed RNA polymerase, subunit A'' (rpoA2)	56.7%
AF1887	DNA-directed RNA polymerase, subunit B' (rpoB1)	65.3%
AF1886	DNA-directed RNA polymerase, subunit B'' (rpoB2)	57.1%
AF2282	DNA-directed RNA polymerase, subunit D (rpoD)	34.6%
AF1117	DNA-directed RNA polymerase, subunit E' (rpoE1)	48.4%
AF1116	DNA-directed RNA polymerase, subunit E'' (rpoE2)	40.0%
AF1885	DNA-directed RNA polymerase, subunit H (rpoH)	59.5%
AF1131	DNA-directed RNA polymerase, subunit K (rpoK)	61.5%
AF0207	DNA-directed RNA polymerase, subunit L (rpoL)	42.0%
AF1130	DNA-directed RNA polymerase, subunit N (rpoN)	58.8%

Transcription factors

AF1813	TBP-interacting protein TIP49	45.7%
AF1299	transcription initiation factor IIB	60.4%
AF0373	transcription initiation factor IID	59.4%
AF0757	transcription initiation factor IIE, subunit alpha, putative	23.5%
AF1891	transcription termination-antitermination factor NusA, putative	48.9%
AF1235	transcription-associated protein TFIIS	59.0%

RNA processing

AF1783	dimethyladenosine transferase (ksgA)	44.7%
AF2067	fibrillarin (fib)	49.3%
AF0482	mRNA 3'-end processing factor, putative	55.5%
AF0532	mRNA 3'-end processing factor, putative	39.1%
AF2361	mRNA 3'-end processing factor, putative	30.5%
AF2399	rRNA methylase, putative	36.4%
AF0362	snRNP, putative	32.0%
AF0875	snRNP, putative	35.7%

TRANSLATION

Amino acyl tRNA synthetases

AF2255	alanyl-tRNA synthetase (alaS)	47.1%
AF0894	arginyl-tRNA synthetase (argS)	48.8%
AF0820	aspartyl-tRNA synthetase (aspS)	62.5%
AF0411	cysteinyl-tRNA synthetase (cysS)	46.1%
AF0260	glutamyl-tRNA synthetase (glX)	44.9%
AF0916	glycyl-tRNA synthetase (glyS)	51.2%
AF1642	histidyl-tRNA synthetase (hisS)	46.0%
AF0633	isoleucyl-tRNA synthetase (ileS)	48.9%
AF2421	leucyl-tRNA synthetase (leuS)	49.7%
AF1216	lysyl-tRNA synthetase (lysS)	43.6%
AF1453	methionyl-tRNA synthetase (metS)	45.2%
AF1955	phenylalanyl-tRNA synthetase, subunit alpha (pheS)	44.4%
AF1424	phenylalanyl-tRNA synthetase, subunit beta (pheT)	42.6%
AF1609	prolyl-tRNA synthetase (proS)	56.8%
AF2035	seryl-tRNA synthetase (serS)	45.4%
AF0648	threonyl-tRNA synthetase (thrS)	46.9%
AF1694	tryptophanyl-tRNA synthetase (trpS)	52.4%
AF0776	tyrosyl-tRNA synthetase (tyrS)	57.6%
AF2224	valyl-tRNA synthetase (valS)	54.5%

Degradation of proteins, peptides, and glycopeptides

AF1976	26S protease regulatory subunit 4	66.0%
AF1663	alkaline serine protease (aprM)	44.5%
AF0678	aminopeptidase, putative	27.8%
AF0364	ATP-dependent protease La (lon)	36.6%
AF1946	cysteine proteinase, putative	36.2%
AF1281	intracellular protease (pfpI)	66.0%
AF1112	O-sialoglycoprotein endopeptidase (gcp)	57.6%
AF0665	O-sialoglycoprotein endopeptidase, putative	35.6%
AF2086	protease inhibitor, putative	37.0%
AF0490	proteasome, subunit alpha (psmA)	60.8%
AF0481	proteasome, subunit beta (psmB)	58.3%
AF2034	X-pro aminopeptidase (pepQ)	34.6%

Protein modification

AF0656	antibiotic maturation protein (pmbA)	32.7%
AF0378	CODH nickel-insertion accessory protein (cooC-1)	35.7%
AF1685	CODH nickel-insertion accessory protein (cooC-2)	47.4%
AF1615	cofactor modifying protein (cmo)	27.2%
AF2196	deoxyhypusine synthase (dys1-1)	32.8%
AF2300	deoxyhypusine synthase (dys1-2)	34.9%
AF0381	diphthine synthase (dph5)	40.8%
AF2324	fmu and fmv protein	40.0%
AF1367	hydrogenase expression /formation protein (hypA)	40.4%
AF1368	hydrogenase expression/formation protein (hypB)	54.4%
AF1369	hydrogenase expression/formation protein (hypC)	40.5%
AF1370	hydrogenase expression/formation protein (hypD)	46.0%
AF1365	hydrogenase expression/formation protein (hypE)	51.9%
AF1366	hydrogenase expression/formation regulatory protein (hypF)	45.1%
AF0036	L-isoaspartyl protein carboxyl methyltransferase (pcm-1)	60.7%
AF2322	L-isoaspartyl protein carboxyl methyltransferase (pcm-2)	59.3%
AF1840	methionyl aminopeptidase (map)	48.6%
AF1989	peptidyl-prolyl cis-trans isomerase (slyD)	34.4%
AF0853	proliferating-cell nucleolar antigen P120, putative	35.7%
AF2039	proliferating-cell nucleolar antigen P120, putative	44.2%
AF1449	pyruvate formate-lyase 2 (pflD)	37.8%
AF1450	pyruvate formate-lyase 2 activating enzyme (pflC)	38.8%
AF0117	pyruvate formate-lyase activating enzyme (act-1)	25.5%
AF0818	pyruvate formate-lyase activating enzyme (act-2)	42.3%
AF1330	pyruvate formate-lyase activating enzyme (act-3)	45.8%
AF2278	pyruvate formate-lyase activating enzyme (act-4)	42.5%
AF1961	pyruvate formate-lyase activating enzyme (pflX)	50.2%
AF0380	transmembrane oligosaccharyl transferase, putative	27.8%
AF0329	transmembrane oligosaccharyl transferase, putative	29.3%

Ribosomal proteins: synthesis and modification

AF1480	LSU ribosomal protein L1P (rpl1P)	48.6%
AF1922	LSU ribosomal protein L2P (rpl2P)	60.4%
AF1925	LSU ribosomal protein L3P (rpl3P)	56.5%
AF1924	LSU ribosomal protein L4P (rpl4P)	56.4%
AF1912	LSU ribosomal protein L5P (rpl5P)	51.7%
AF1909	LSU ribosomal protein L6P (rpl6P)	53.7%
AF0764	LSU ribosomal protein L7AE (rpl7AE)	60.7%
AF1491	LSU ribosomal protein L10E (rpl10E)	45.6%
AF0538	LSU ribosomal protein L11P (rpl11P)	67.8%
AF1482	LSU ribosomal protein L12A (rpl12A)	76.0%
AF1128	LSU ribosomal protein L13P (rpl13P)	47.4%
AF1915	LSU ribosomal protein L14P (rpl14P)	66.7%
AF2319	LSU ribosomal protein L15E (rpl15E)	70.3%
AF1903	LSU ribosomal protein L15P (rpl15P)	53.8%
AF1127	LSU ribosomal protein L18E (rpl18E)	53.8%
AF1906	LSU ribosomal protein L18P (rpl18P)	57.8%
AF1907	LSU ribosomal protein L19E (rpl19E)	55.5%
AF1529	LSU ribosomal protein L21E (rpl21E)	53.2%
AF1920	LSU ribosomal protein L22P (rpl22P)	55.2%
AF1923	LSU ribosomal protein L23P (rpl23P)	55.8%
AF0537	LSU ribosomal protein L24A (rpl24A)	51.4%
AF0766	LSU ribosomal protein L24E (rpl24E)	66.1%
AF1914	LSU ribosomal protein L24P (rpl24P)	57.8%
AF1918	LSU ribosomal protein L29P (rpl29P)	44.8%
AF1890	LSU ribosomal protein L30E (rpl30E)	41.7%
AF1904	LSU ribosomal protein L30P (rpl30P)	55.9%
AF2066	LSU ribosomal protein L31E (rpl31E)	50.6%
AF1908	LSU ribosomal protein L32E (rpl32E)	51.2%
AF0057	LSU ribosomal protein L37AE (rpl37AE)	47.8%
AF0874	LSU ribosomal protein L37E (rpl37E)	57.9%
AF2067	LSU ribosomal protein L39E (rpl39E)	56.9%
AF1430	LSU ribosomal protein L40E (rpl40E)	73.3%
AF1333	LSU ribosomal protein L44E (rpl44E)	46.8%
AF2064	LSU ribosomal protein LXA (rplXA)	53.8%
AF0739	ribosomal protein S18 alanine acetyltransferase	38.5%
AF2303	ribosomal protein S6 modification protein (rimK)	32.2%
AF1133	SSU ribosomal protein S2P (rps2P)	58.3%
AF1919	SSU ribosomal protein S3P (rps3P)	50.0%
AF1913	SSU ribosomal protein S4E (rps4E)	48.9%
AF2284	SSU ribosomal protein S4P (rps4P)	59.1%
AF1905	SSU ribosomal protein S5P (rps5P)	60.0%
AF0511	SSU ribosomal protein S6E (rps6E)	50.8%
AF1893	SSU ribosomal protein S7P (rps7P)	59.6%
AF2152	SSU ribosomal protein S8E (rps8E)	61.6%
AF1910	SSU ribosomal protein S8P (rps8E)	64.6%
AF1129	SSU ribosomal protein S9P (rps9P)	59.5%
AF0938	SSU ribosomal protein S10P (rps10P)	71.0%
AF2283	SSU ribosomal protein S11P (rps11P)	71.1%
AF1892	SSU ribosomal protein S12P (rps12P)	74.1%
AF2285	SSU ribosomal protein S13P (rps13P)	52.1%
AF1911	SSU ribosomal protein S14P (rps14P)	61.5%
AF0801	SSU ribosomal protein S15P (rps15P)	62.0%
AF0911	SSU ribosomal protein S17E (rps17E)	52.6%
AF1916	SSU ribosomal protein S17P (rps17P)	59.0%
AF2069	SSU ribosomal protein S19E (rps19E)	64.2%
AF1921	SSU ribosomal protein S19P (rps19P)	60.9%
AF1114	SSU ribosomal protein S24E (rps24E)	40.2%
AF1113	SSU ribosomal protein S27AE (rps27AE)	60.0%
AF1334	SSU ribosomal protein S27E (rps27E)	49.0%
AF0785	SSU ribosomal protein S28E (rps28E)	55.6%
AF2320	SSU ribosomal protein S3AE (rps3AE)	38.9%

tRNA modification

AF0588	archaeosine tRNA-ribosyltransferase (tgtA)	52.0%
AF1954	Glu-tRNA amidotransferase, subunit A (gatA-1)	38.6%
AF2328	Glu-tRNA amidotransferase, subunit A (gatA-2)	53.5%
AF1440	Glu-tRNA amidotransferase, subunit B (gatB-1)	54.7%
AF2116	Glu-tRNA amidotransferase, subunit B (gatB-2)	46.4%

AF2328	Glu-tRNA amidotransferase, subunit C (gatC)	35.1%
AF0815	N2,N2-dimethylguanosine tRNA methyltransferase (trm1)	38.2%
AF1730	pseudouridylate synthase I (truA)	37.4%
AF1485	queuine tRNA-ribosyltransferase (tgtB)	44.1%
AF0493	ribonuclease PH (rph)	30.8%
AF0900	tRNA intron endonuclease (endA)	41.8%
AF2156	tRNA nucleotidyltransferase (cca)	43.9%
Translation factors		
AF2350	ATP-dependent RNA helicase HepA, putative	31.5%
AF2254	ATP-dependent RNA helicase, DEAD-family (deaD)	52.2%
AF0071	ATP-dependent RNA helicase, putative	29.6%
AF1468	ATP-dependent RNA helicase, putative	48.1%
AF2406	ATP-dependent RNA helicase, putative	35.2%
AF1149	large helicase-related protein (lhr-1)	34.5%
AF2177	large helicase-related protein (lhr-2), authentic frameshift	56.0%
AF1220	peptide chain release factor eRF, subunit 1	51.2%
AF2245	SKI2-family helicase, authentic frameshift	45.7%
AF0937	translation elongation factor EF-1, subunit alpha (tuf)	74.4%
AF0574	translation elongation factor EF-1, subunit beta	31.3%
AF1894	translation elongation factor EF-2 (fus)	62.5%
AF0777	translation initiation factor eIF-1A (eif1A)	57.5%
AF0627	translation initiation factor eIF-2, subunit alpha (eif2A)	51.1%
AF2326	translation initiation factor eIF-2, subunit beta, putative	46.5%
AF0592	translation initiation factor eIF-2, subunit gamma (eif2G)	64.4%
AF0370	translation initiation factor eIF-2B, subunit delta (eif2BD)	53.3%
AF2037	translation initiation factor eIF-2B, subunit delta (eif2BD)	57.9%
AF0645	translation initiation factor eIF-5A (eif5A)	50.4%
AF0768	translation initiation factor IF-2 (infB)	52.2%

TRANSPORT AND BINDING PROTEINS

General		
AF0393	ABC transporter, ATP-binding protein	34.5%
AF0984	ABC transporter, ATP-binding protein	35.2%
AF1006	ABC transporter, ATP-binding protein	35.1%
AF1018	ABC transporter, ATP-binding protein	57.7%
AF1021	ABC transporter, ATP-binding protein	37.8%
AF1136	ABC transporter, ATP-binding protein	39.3%
AF1139	ABC transporter, ATP-binding protein	38.2%
AF1300	ABC transporter, ATP-binding protein	34.1%
AF1469	ABC transporter, ATP-binding protein	43.5%
AF1819	ABC transporter, ATP-binding protein	51.1%
AF1982	ABC transporter, ATP-binding protein	41.3%
AF2364	ABC transporter, ATP-binding protein	53.5%
AF1006	ABC transporter, ATP-binding protein, putative	28.7%
AF1064	ABC transporter, ATP-binding protein, putative	36.0%
AF1983	ABC transporter, periplasmic binding protein	25.4%
AF1981	ABC transporter, permease protein	29.9%
AF1995	sodium- and chloride-dependent transporter	52.5%
Amino acids, peptides and amines		
AF1766	amino-acid ABC transporter, periplasmic binding protein/protein kinase	27.4%
AF0222	branched-chain amino acid ABC transporter, ATP-binding protein (braF-1)	42.7%
AF0822	branched-chain amino acid ABC transporter, ATP-binding protein (braF-2)	44.7%
AF0959	branched-chain amino acid ABC transporter, ATP-binding protein (braF-3)	37.6%
AF1390	branched-chain amino acid ABC transporter, ATP-binding protein (braF-4)	59.7%
AF0221	branched-chain amino acid ABC transporter, ATP-binding protein (braG-1)	48.2%
AF0823	branched-chain amino acid ABC transporter, ATP-binding protein (braG-2)	42.9%
AF0958	branched-chain amino acid ABC transporter, ATP-binding protein (braG-3)	34.1%
AF1389	branched-chain amino acid ABC transporter, ATP-binding protein (braG-4)	64.6%
AF0223	branched-chain amino acid ABC transporter, periplasmic binding protein (braC-1)	34.3%
AF0827	branched-chain amino acid ABC transporter, periplasmic binding protein (braC-2)	26.8%
AF0962	branched-chain amino acid ABC transporter, periplasmic binding protein (braC-3)	25.6%
AF1391	branched-chain amino acid ABC transporter, periplasmic binding protein (braC-4)	50.1%
AF0224	branched-chain amino acid ABC transporter, permease protein (braD-1)	25.4%
AF0825	branched-chain amino acid ABC transporter, permease protein (braD-2)	30.8%
AF0961	branched-chain amino acid ABC transporter, permease protein (braD-3)	23.9%
AF1392	branched-chain amino acid ABC transporter, permease protein (braD-4)	65.4%
AF0225	branched-chain amino acid ABC transporter, permease protein (braE-1)	28.7%
AF0824	branched-chain amino acid ABC transporter, permease protein (braE-2)	31.3%
AF0960	branched-chain amino acid ABC transporter, permease protein (braE-3)	30.1%
AF1393	branched-chain amino acid ABC transporter, permease protein (braE-4)	60.5%
AF1612	cationic amino acid transporter (cat-1)	29.5%
AF1774	cationic amino acid transporter (cat-2)	38.0%
AF1770	dipeptide ABC transporter, ATP-binding protein (dppD)	47.8%
AF1771	dipeptide ABC transporter, ATP-binding protein (dppF)	43.1%
AF1767	dipeptide ABC transporter, dipeptide-binding protein (dppA)	33.1%
AF1768	dipeptide ABC transporter, permease protein (dppB)	39.3%
AF1769	dipeptide ABC transporter, permease protein (dppC)	40.8%
AF0680	glutamine ABC transporter, ATP-binding protein (glnQ)	63.8%
AF0231	glutamine ABC transporter, periplasmic glutamine-binding protein (glnH)	38.0%
AF0232	glutamine ABC transporter, permease protein (glnP)	39.3%
AF0881	osmoprotection protein (proV)	39.0%
AF0979	osmoprotection protein (proW-1)	32.8%
AF0980	osmoprotection protein (proW-2)	36.8%
AF0982	osmoprotection protein (proX)	28.7%
AF0015	proline permease (putP-1)	26.2%
AF0869	proline permease (putP-2)	27.4%
AF1222	proline permease (putP-3)	27.0%
AF1608	spermidine/putrescine ABC transporter, ATP-binding protein (potA)	50.2%
AF1606	spermidine/putrescine ABC transporter, periplasmic spermidine/putrescine-binding protein (potD), authentic frameshift	31.0%
AF1607	spermidine/putrescine ABC transporter, permease protein (potB)	38.0%
AF1605	spermidine/putrescine ABC transporter, permease protein (potC)	38.7%
Anions		
AF2308	arsenite transport protein (arsB)	27.3%
AF1415	chloride channel, putative	27.2%
AF0025	cyanate transport protein (cynX)	24.5%
AF0087	nitrate ABC transporter, ATP-binding protein (nrtC-1)	47.4%
AF0638	nitrate ABC transporter, ATP-binding protein (nrtC-2)	55.5%
AF0640	nitrate ABC transporter, ATP-binding protein, putative	32.5%
AF0086	nitrate ABC transporter, permease protein (nrtB-1)	35.4%
AF0639	nitrate ABC transporter, permease protein (nrtB-2)	37.4%
AF1359	phosphate ABC transporter, ATP-binding protein (pstB)	66.0%
AF1356	phosphate ABC transporter, periplasmic phosphate-binding protein (phoX)	25.1%
AF1358	phosphate ABC transporter, permease protein (pstA)	34.1%
AF1357	phosphate ABC transporter, permease protein (pstC)	33.7%
AF1360	phosphate ABC transporter, regulatory protein (phoU)	26.9%
AF0791	phosphate permease, putative	31.1%
AF1798	phosphate permease, putative	62.9%
AF0092	sulfate ABC transporter, ATP-binding protein (cysA)	54.2%
AF0093	sulfate ABC transporter, permease protein (cysT)	44.1%
Carbohydrates, organic alcohols, and acids		
AF0347	C4-dicarboxylate transporter (maeI)	24.5%
AF1426	glycerol uptake facilitator, MIP channel (glpF)	36.2%
AF0013	hexuronate transporter (exuT)	25.1%
AF0806	L-lactate permease (lctP)	31.7%
AF0008	oxalate/formate antiporter (oxfT-1)	25.7%
AF0367	oxalate/formate antiporter (oxfT-2)	33.2%
AF1069	pantothenate permease (panF-1)	28.9%
AF1205	pantothenate permease (panF-2)	24.8%
AF0237	pantothenate permease (panF-3)	25.1%
AF0041	polysaccharide ABC transporter, ATP-binding protein (rfbB-1)	42.5%
AF0290	polysaccharide ABC transporter, ATP-binding protein (rfbB-2)	43.9%
AF0042	polysaccharide ABC transporter, permease protein (rfbA-1)	27.5%
AF0289	polysaccharide ABC transporter, permease protein (rfbA-2)	28.5%
AF0887	ribose ABC transporter, ATP-binding protein (rbsA-1)	33.3%
AF1170	ribose ABC transporter, ATP-binding protein (rbsA-2)	27.9%
AF0888	ribose ABC transporter, permease protein (rbsC-1)	24.1%
AF0889	ribose ABC transporter, permease protein (rbsC-2)	31.2%
AF2014	sugar transporter, putative	26.0%
Cations		
AF0977	ammonium transporter (amt-1)	44.3%
AF1746	ammonium transporter (amt-2)	49.0%
AF1749	ammonium transporter (amt-3)	41.5%
AF0473	cation-transporting ATPase, P-type (pacS)	44.0%
AF0152	copper-transporting ATPase, P-type (copB)	44.5%
AF0246	iron (II) transporter (feoB-1)	33.3%
AF2394	iron (II) transporter (feoB-2)	48.0%
AF0661	iron (II) transporter (feoB-3), authentic frameshift	29.4%
AF0430	iron (III) ABC transporter, ATP-binding protein (hemV-1)	50.4%
AF0432	iron (III) ABC transporter, ATP-binding protein (hemV-2)	58.7%
AF1401	iron (III) ABC transporter, ATP-binding protein (hemV-3)	35.2%
AF1397	iron (III) ABC transporter, periplasmic hemin-binding protein (hemT), authentic frameshift	28.2%
AF0431	iron (III) ABC transporter, permease protein (hemU-1)	36.2%
AF1402	iron (III) ABC transporter, permease protein (hemU-2)	35.2%
AF0786	magnesium and cobalt transporter (corA)	40.1%
AF0346	mercuric transport protein periplasmic component (merP)	35.2%
AF0217	Na+/H+ antiporter (napA-1)	28.2%
AF1245	Na+/H+ antiporter (napA-2)	28.4%
AF0848	Na+/H+ antiporter (nhe2)	33.1%
AF0715	potassium channel, putative	39.5%
AF1673	potassium channel, putative	36.3%
AF2197	potassium channel, putative	24.6%
AF0218	TRK potassium uptake system protein (trkA-1)	30.2%
AF0638	TRK potassium uptake system protein (trkA-2)	42.9%
AF0839	TRK potassium uptake system protein (trkH)	39.8%
Other		
AF0834	ferritin, putative	39.8%
AF1980	heme exporter protein C (helC)	29.0%
AF1144	multidrug resistance protein	29.2%
AF1325	multidrug resistance protein	29.9%
AF2258	multidrug resistance protein	31.3%

OTHER CATEGORIES

Adaptations and atypical conditions		
AF0508	ethylene-inducible protein	74.5%
AF0235	heat shock protein (htpX)	32.9%
AF0942	surE stationary-phase survival protein (surE)	50.2%
AF1996	virulence associated protein C (vapC-1)	50.0%
AF1690	virulence associated protein C (vapC-2)	30.0%
Drug and analog sensitivity		
AF1884	daunorubicin resistance ATP-binding protein (drrA)	47.1%
AF1883	daunorubicin resistance membrane protein (drrB)	27.0%
AF0487	penicillin G acylase	31.7%
AF1214	phenylacrylic acid decarboxylase (pad1)	43.2%
AF2194	rRNA (adenine-N6)-methyltransferase, putative	29.2%
AF1696	small multidrug export protein (qacE)	39.0%
Transposon-related functions		
AF0120	insertion sequence ISH S1, authentic frameshift	34.5%
AF0193	ISA0963-1, putative transposase, authentic frameshift	34.3%
AF0309	ISA0963-2, putative transposase	33.5%
AF1310	ISA0963-3, putative transposase	33.5%
AF1383	ISA0963-4, putative transposase	33.5%
AF1410	ISA0963-5, putative transposase	33.5%
AF1705	ISA0963-6, putative transposase	33.5%
AF1836	ISA0963-7, putative transposase, authentic frameshift	20.0%
AF0678	ISA1083-1, ISORF2	33.6%
AF0679	ISA1083-1, putative transposase	37.2%
AF1351	ISA1083-2, ISORF2	30.8%
AF1352	ISA1083-2, putative transposase	31.5%
AF2140	ISA1083-3, ISORF2	30.8%
AF2139	ISA1083-3, putative transposase	31.5%
AF0278	ISA1214-1, ISORF2	27.7%
AF0279	ISA1214-1, putative transposase	33.3%
AF0305	ISA1214-2, ISORF2	27.7%
AF0306	ISA1214-2, putative transposase	33.3%
AF0641	ISA1214-3, ISORF2	26.5%
AF0642	ISA1214-3, putative transposase	33.3%
AF0857	ISA1214-4, ISORF2	27.7%
AF0858	ISA1214-4, putative transposase	33.3%
AF2091	ISA1214-5, ISORF2	26.5%
AF2092	ISA1214-5, putative transposase	33.3%
AF2223	ISA1214-6, ISORF2	26.5%
AF2222	ISA1214-6, putative transposase	25.6%
AF0138	transposase IS240-A	43.3%
AF0895	transposase IS240-A	46.2%
AF2390	transposase, authentic frameshift	24.0%
AF0137	transposase, putative	29.6%
AF1628	transposase, putative	32.8%

UNKNOWN

AF0477	AAA superfamily ATPase	35.0%
AF0613	allene oxide synthase, putative	39.5%
AF0478	ATP-binding protein PhnP (phnP)	30.9%
AF1775	atrazine chlorohydrolase, putative	34.4%
AF0973	bile acid-inducible operon protein F (baiF-1)	30.8%
AF0974	bile acid-inducible operon protein F (baiF-2)	29.9%
AF1315	bile acid-inducible operon protein F (baiF-3)	31.3%
AF2063	c-myc binding protein, putative	21.7%
AF1992	calcium-binding protein, putative	31.2%
AF2287	carotenoid biosynthetic gene ERWCRTS, putative	49.4%
AF0512	chloroplast inner envelope membrane protein	42.5%
AF2251	competence-damage protein, putative	26.0%
AF0090	dehydrase	34.1%
AF1498	dehydrase, putative	29.4%
AF1518	DNA/pantothenate metabolism flavoprotein, putative	51.4%
AF0039	dolichol-P-glucose synthetase, putative	33.7%
AF0328	dolichol-P-glucose synthetase, putative	39.0%
AF0581	dolichol-P-glucose synthetase, putative	27.5%
AF0569	DR-beta chain MHC class II	37.7%
AF0383	endonuclease III, putative	47.1%
AF1150	erpK protein, putative	54.9%
AF2372	extragenic suppressor (suhB)	37.0%
AF1418	glycerol-3-phosphate cytidyltransferase (tagD)	56.6%
AF0744	GTP-binding protein	33.4%
AF1181	GTP-binding protein	36.3%
AF1364	GTP-binding protein	57.5%
AF2146	GTP-binding protein	65.9%
AF0428	GTP-binding protein, GTP1/OBG-family	43.9%
AF2237	HAM1 protein	31.4%
AF2211	HIT family protein (hit)	29.6%
AF0216	L-isoaspartyl protein carboxyl methyltransferase PimT, putative	35.5%
AF2313	macC protein (macC)	43.0%
AF0429	methyltransferase	43.8%
AF0186	nifS protein, class-V aminotransferase (nifS-1)	46.1%
AF0564	nifS protein, class-V aminotransferase (nifS-2)	45.1%
AF0185	nifU protein (nifU-1)	55.6%
AF0565	nifU protein (nifU-2)	55.6%
AF0632	nifU protein (nifU-3)	47.4%
AF1781	nodulation protein NfeD (nfeD)	33.4%
AF2269	nucleotide-binding protein	48.7%
AF2382	nucleotide-binding protein	49.1%
AF0374	p-nitrophenyl phosphatase (pho2)	31.7%
AF1978	periplasmic divalent cation tolerance protein (cutA)	31.3%
AF1662	prepro-subtilisin sendai, putative	35.6%
AF2021	rod shape-determining protein (mreB)	26.6%
AF1778	stage V sporulation protein (spoVG)	43.9%
AF1970	TPR domain-containing protein	29.0%
AF2202	tryptophan-specific permease, putative	25.2%
AF0816	vtpJ-therm, putative	42.1%
AF1679	vtpJ-therm, putative	45.1%

NiceProt View of TrEMBL: O29897

Note: most headings are clickable, even if they don't appear as links. They link to the user manual or other documents.

Entry information

Entry name	**O29897**
Primary accession number	**O29897**
Secondary accession numbers	None
Entered in TrEMBL in	Release 05, January 1998
Sequence was last modified in	Release 05, January 1998
Annotations were last modified in	Release 24, June 2003

Name and origin of the protein

Protein name	**Conserved hypothetical transmembrane protein**
Synonyms	None
Gene name	**AF0350**
From	Archaeoglobus fulgidus [TaxID: 2234]
Taxonomy	Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales; Archaeoglobaceae; Archaeoglobus.

References

[1] SEQUENCE FROM NUCLEIC ACID.
STRAIN=VC-16 / DSM 4304 / ATCC 49558;
MEDLINE=98049343; PubMed=9389475; [NCBI, ExPASy, EBI, Israel, Japan]
Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E., Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D., Richardson D.L., Kerlavage A.R., Graham D.E., Kyrpides N.C., Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S., Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B., Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L., Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T., Cotton M.D., Spriggs T., Artiach P., Kaine B.P., Sykes S.M., Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A., Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
"The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon Archaeoglobus fulgidus.";
Nature 390:364-370(1997).

Comments

None

Cross-references

EMBL	AE001080; AAB90884.1; -.[EMBL / GenBank / DDBJ] [CoDingSequence]
TIGR	AF0350; -.
GO	GO:0016021;Cellular component: integral to membrane *(inferred from electronic annotation).*
ProDom	[Domain structure / List of seq. sharing at least 1 domain]
HOBACGEN	[Family / Alignment / Tree]
ProtoMap	O29897.
PRESAGE	O29897.
ModBase	O29897.
SMR	O29897; 927825BC1EFC1B46.
SWISS-2DPAGE	Get region on 2D PAGE.
UniRef	View cluster of proteins with at least 50% / 90% identity.

Keywords

Hypothetical protein; **Transmembrane**; **Complete proteome**.

Features

None

Sequence information

Length: **211 AA** Molecular weight: **23303 Da** CRC64: **927825BC1EFC1B46** [This is a checksum on the sequence]

```
         10          20          30          40          50          60
          |           |           |           |           |           |
MNGPDNRHAA FRYRADSRNA RRSAWHWWLR YNAPCPRIPL QLPAAGSDRD DDNGCDFNRK

         70          80          90         100         110         120
          |           |           |           |           |           |
LRSCRPHKAE ERGLLNCKGC CGKRCGWGCS WEHNLLLHRK PAVASELHPR LCLPLRQHRM

        130         140         150         160         170         180
          |           |           |           |           |           |
IYEGTIKRAM PEKPGNEVPG SSTAKATIGF VIGIITGIVG LGGGYALVPS FIYLLGSAVK

        190         200         210
          |           |           |
IAVGTSLAEV LLPPFLPQTK VHMEGNQKGS L
```

O29897 in FASTA format

View entry in original TrEMBL format
View entry in raw text format (no links)
Request for annotation of this TrEMBL entry

BLAST BLAST submission on ExPASy/SIB or at NCBI (USA)

Sequence analysis tools: ProtParam, ProtScale, Compute pI/Mw, PeptideMass, PeptideCutter, Dotlet (Java)

ScanProsite, MotifScan

Search the SWISS-MODEL Repository

ExPASy Home page | Site Map | Search ExPASy | Contact us | Swiss-Prot

Hosted by YPRC Korea | Mirror sites: | Australia | Bolivia | Canada | China | Switzerland | Taiwan | USA

10. South, T. L., Blake, P. R., Hare, D. R. & Summers, M. F. C-terminal retroviral-type zinc finger domain from the HIV-1 nucleocapsid protein is structurally similar to the N-terminal zinc finger domain. *Biochemistry* **30**, 6342–6349 (1991).
11. Luisi, B. F. *et al.* Crystallographic analysis of the interaction of the glucorticoid receptor with DNA. *Nature* **352**, 497–505 (1991).
12. Marmorstein, R., Carey, M., Ptashne, M. & Harrison, S. C. DNA recognition by GAL4: structure of a protein–DNA complex. *Nature* **356**, 408–414 (1992).
13. Everett, R. D. *et al.* A novel arrangement of zinc-binding residues and secondary structure in the C3HC4 motif of an alpha herpes virus protein family. *J. Mol. Biol.* **234**, 1038–1047 (1993).
14. Barlow, P. N., Luisi, B., Milner, A., Elliot, M. & Everett, R. Structure of the C_3HC_4 domain by ^{1}H-nuclear magnetic resonance spectroscopy. *J. Mol. Biol.* **237**, 201–211 (1994).
15. Borden, K. L. B. *et al.* The solution structure of the RING finger domain from the acute promyelocytic leukaemia proto-oncoprotein PML. *EMBO J.* **14**, 1532–1541 (1995).
16. Phillips, S. E. V. The β-ribbon DNA recognition motif. *Ann. Rev. Biophys. Biomol. Struct.* **23**, 671–701 (1994).
17. Kim, J. L., Nikolov, D. B. & Burley, S. K. Co-crystal structure of TBP recognizing the minor groove of a TATA element. *Nature* **365**, 520–527 (1993).
18. Kim, Y., Geiger, J. H., Hahn, S. & Sigler, P. B. Crystal structure of a yeast TBP–TATA-box complex. *Nature* **365**, 512–520 (1993).
19. Schumacher, M. A., Choi, K. Y., Zalkin, H. & Brennan, R. G. Crystal structure of LacI member, PurR, bound to DNA: minor groove binding by α-helices. *Science* **266**, 763–770 (1994).
20. Flick, K. E. *et al.* Crystallization and preliminary X-ray studies of I-PpoI: a nuclear, intron-encoded homing endonuclease from *Physarum polycephalum*. *Protein Sci.* **6**, 1–4 (1997).
21. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
22. Leslie, A. G. W. in *Joint CCP4 and ESF-EACMB Newsletter on Protein Crystallography* (Daresbury Laboratory, Warrington, UK, 1992).
23. CCP4 *The SERC (UK) Collaborative Computing Project No. 4, a Suite of Programs for Protein Crystallography* (Daresbury Laboratory, Warrington, UK, 1979).
24. QUANTA96 *X-ray Structure Analysis User's Reference* (Molecular Simulations, San Diego, 1996).
25. Brünger, A. *XPLOR version 3.1: A System for X-ray Crystallography and NMR* (Yale Univ. Press, New Haven, CT, 1992).
26. Laskowski, R. J., Macarthur, M. W., Moss, D. S. & Thornton, J. M. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 383–290 (1993).
27. Evans, S. V. SETOR: hardware-lighted three-dimensional solid model representations of macromolecules. *J. Mol. Graphics* **11**, 134–138 (1993).

Acknowledgements. We thank D. McHugh, K. Stephens and J. D. Heath for initial subcloning, purification and crystallization studies; R. Strong, K. Zhang and B. Scott for advice during the crystallographic analysis; and the beamline staff at the Advanced Light Source (NLBL laboratories), beamline 5.0.2, particularly T. Earnest, for assistance. B.L.S. and R.J.M. are funded for this project by the NIH. K.E.F. was supported by an NIH training grant and the American Heart Associaiton. M.S.J. was supported by an NSF fellowship and an NIH training grant.

Correspondence and requests for materials and coordinates should be addressed to B.L.S. (e-mail: bstoddar@fred.fhcrc.org). Coordinates have been deposited in the Brookhaven Protein Data Bank (accession nos lipp, 1a73, 1a74).

corrections

Emergence of symbiosis in peptide self-replication through a hypercyclic network

David H. Lee, Kay Severin, Yohei Yokobayashi & M. Reza Ghadiri

Nature **390**, 591–594 (1997)

Hypercycles are based on second-order (or higher) autocatalysis and defined by two or more replicators that are connected by another superimposed autocatalytic cycle. Our study describes a mutualistic relationship between two replicators, each catalysing the formation of the other, that are linked by a superimposed catalytic cycle. Although the kinetic data suggest the intermediary of higher-order species in the autocatalytic processes, the present system should not be referred to as an example of a minimal hypercycle in the absence of direct experimental evidence for the autocatalytic cross-coupling between replicators. □

The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon *Archaeoglobus fulgidus*

Hans-Peter Klenk, Rebecca A. Clayton, Jean-Francois Tomb, Owen White, Karen E. Nelson, Karen A. Ketchum, Robert J. Dodson, Michelle Gwinn, Erin K. Hickey, Jeremy D. Peterson, Delwood L. Richardson, Anthony R. Kerlavage, David E. Graham, Nikos C. Kyrpides, Robert D. Fleischmann, John Quackenbush, Norman H. Lee, Granger G. Sutton, Steven Gill, Ewen F. Kirkness, Brian A. Dougherty, Keith McKenney, Mark D. Adams, Brendan Loftus, Scott Peterson, Claudia I. Reich, Leslie K. McNeil, Jonathan H. Badger, Anna Glodek, Lixin Zhou, Ross Overbeek, Jeannine D. Gocayne, Janice F. Weidman, Lisa McDonald, Teresa Utterback, Matthew D. Cotton, Tracy Spriggs, Patricia Artiach, Brian P. Kaine, Sean M. Sykes, Paul W. Sadow, Kurt P. D'Andrea, Cheryl Bowman, Claire Fujii, Stacey A. Garland, Tanya M. Mason, Gary J. Olsen, Claire M. Fraser, Hamilton O. Smith, Carl R. Woese & J. Craig Venter

Nature **390**, 364–370 (1997)

The pathway for sulphate reduction is incorrect as published: in Fig. 3 on page 367, adenylyl sulphate 3-phosphotransferase (*cysC*) is not needed in the pathway as outlined, as adenylyl sulphate reductase (*aprAB*) catalyses the first step in the reduction of adenylyl sulphate. The correct sequence of reactions is: sulphate is first activated to adenylyl sulphate, then reduced to sulphite and subsequently to sulphide. The enzymes catalysing these reactions are: sulphate adenylyltransferase (*sat*), adenylylsulphate reductase (*aprAB*), and sulphite reductase (*dsrABD*). We thank Jens-Dirk Schwenn for bringing this error to our attention. □

9. T. Kunik *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 1871 (2001).
10. J. M. Young, L. D. Kuykendall, E. Martinez-Romero, A. Kerr, H. Sawada, *Int. J. Syst. Evol. Microbiol.* **51**, 89 (2001).
11. E. Martinez, R. Palacios, F. Sanchez, *J. Bacteriol.* **169**, 2828 (1987).
12. F. Galibert *et al.*, *Science* **293**, 668 (2001).
13. T. Kaneko *et al.*, *DNA Res.* **7**, 331 (2000).
14. The complete annotated sequence, detailed methods, and all supplementary data, which will be periodically updated, are available at our Web site (www.agrobacterium.org). Supplemental data are also available at *Science* Online (www.sciencemag.org/cgi/content/full/294/5550/2317/DC1). The sequence has been deposited at GenBank (accession numbers: circular chromosome, AE008688; linear chromosome, AE008689; pAtC58, AE008687; pTiC58, AE008690).
15. At a late phase of this project, we became aware of an independent effort by Goodner *et al.* (*52*) to sequence the C58 strain of *A. tumefaciens*. Comparison of the two independently finished sequences indicates excellent overall agreement. There are two small insertions in the other sequence, one in the circular chromosome and one in pAtC58, relative to our sequence. Single nucleotide discrepancies are far below current accuracy standards for finished genomic sequence (99.99%). Further work will determine which of the observed discrepancies reflect strain differences and which reflect sequencing errors.
16. R. H. Hamilton, M. Z. Fall, *Experientia* **27**, 229 (1971).
17. A. Allardet-Servent, S. Michaux-Charachon, E. Jumas-Bilak, L. Karayan, M. Ramuz, *J. Bacteriol.* **175**, 7869 (1993).
18. J. Zhu *et al.*, *J. Bacteriol.* **182**, 3885 (2000).
19. V. S. Kalogeraki, S. C. Winans, *J. Bacteriol.* **180**, 5660 (1998).
20. R. L. Tatusov *et al.*, *Nucleic Acids Res.* **29**, 22 (2001).
21. J. A. Eisen, J. F. Heidelberg, O. White, S. L. Salzberg, *Genome Biol.* **1**, Research 0011.1 (2000).
22. C. M. Fraser *et al.*, *Nature* **390**, 580 (1997).
23. J. N. Volff, J. Altenbuchner, *FEMS Microbiol. Lett.* **186**, 143 (2000).
24. K. Bao, S. N. Cohen, *Genes Dev.* **15**, 1518 (2001).
25. M. A. Ramirez-Romero, N. Soberon, A. Perez-Oseguera, J. Tellez-Sosa, M. A. Cevallos, *J. Bacteriol.* **182**, 3117 (2000).
26. L. Chen, unpublished data.
27. S. K. Farrand, in *The Rhizobiaceae*, H. P. Spaink, A. Kondorosi, P. J. J. Hooykaas, Eds. (Kluwer Academic, Dordrecht, Netherlands, 1998), pp. 199–233.
28. I. T. Paulsen, L. Nguyen, M. K. Sliwinski, R. Rabus, M. H. Saier Jr., *J. Mol. Biol.* **301**, 75 (2000).
29. J. A. Forward, M. C. Behrendt, N. R. Wyborn, R. Cross, D. J. Kelly, *J. Bacteriol.* **179**, 5482 (1997).
30. R. Sanchez-Fernandez, T. G. Davies, J. O. Coleman, P. A. Rea, *J. Biol. Chem.* **276**, 30231 (2001).
31. C. K. Stover *et al.*, *Nature* **406**, 959 (2000).
32. D. Missiakas, S. Raina, *Mol. Microbiol.* **28**, 1059 (1998).
33. G. T. Robertson, R. M. Roop Jr., *Mol. Microbiol.* **34**, 690 (1999).
34. A. G. Matthysse, G. W. Kijne, in *The Rhizobiaceae*, H. P. Spaink, A. Kondorosi, P. J. J. Hooykaas, Eds. (Kluwer Academic, Dordrecht, Netherlands, 1998), pp. 235–249.
35. K. J. Fullner, J. C. Lara, E. W. Nester, *Science* **273**, 1107 (1996).
36. P. L. Li, I. Hwang, H. Miyagi, H. True, S. K. Farrand, *J. Bacteriol.* **181**, 5033 (1999).
37. J. R. Harper, T. J. Silhavy, in *Principles of Bacterial Pathogenesis*, E. A. Groisman, Ed. (Academic Press, San Diego, CA, 2001), pp. 43–74.
38. J. H. Weiner *et al.*, *Cell* **93**, 93 (1998).
39. K. LeVier, R. W. Phillips, V. K. Grippe, R. M. Roop II, G. C. Walker, *Science* **287**, 2492 (2000).
40. M. J. Pallen, B. W. Wren, *Mol. Microbiol.* **26**, 209 (1997).
41. A. Ichige, G. C. Walker, *J. Bacteriol.* **179**, 209 (1997).
42. S. A. Coleman, M. F. Minnick, *Infect. Immun.* **69**, 4373 (2001).
43. M. Carsiotis, B. A. Stocker, D. L. Weinstein, A. D. O'Brien, *Infect. Immun.* **57**, 3276 (1989).
44. Y. Dessaux, A. Petit, S. K. Farrand, P. J. Murphy, in *The Rhizobiaceae*, H. P. Spaink, A. Kondorosi, P. J. J. Hooykaas, Eds. (Kluwer Academic, Dordrecht, Netherlands, 1998), pp. 173–197.
45. V. S. Kalogeraki, J. Zhu, A. Eberhard, E. L. Madsen, S. C. Winans, *Mol. Microbiol.* **34**, 512 (1999).
46. K. J. Miller, J. M. Wood, *Annu. Rev. Microbiol.* **50**, 101 (1996).
47. Z. Q. Luo, S. K. Farrand, *J. Bacteriol.* **181**, 618 (1999).
48. B. Minambres, E. R. Olivera, R. A. Jensen, J. M. Luengo, *J. Biol. Chem.* **275**, 39529 (2000).
49. E. Jumas-Bilak, S. Michaux-Charachon, G. Bourg, M. Ramuz, A. Allardet-Servent, *J. Bacteriol.* **180**, 2749 (1998).
50. H. Ochman, N. A. Moran, *Science* **292**, 1096 (2001).
51. J. A. Downie, J. P. Young, *Nature* **412**, 597 (2001).
52. B. Goodner *et al.*, *Science* **294**, 2323 (2001).

53. We thank P. Green for useful discussions, R. Gibson for providing the genome_plot program, and J. Staley for critical review of the manuscript. Supported by NSF grant 0092815 (M.V.O.), NIH Public Health Service grant GM32618 (E.W.N.), NIH doctoral fellowship GM19642 (D.W.W.), Sao Paulo State Research Foundation (FAPESP) sabbatical fellowship 1999/11876-5 (J.C.S.), a FUNDECT-MS grant (N.F.A.), and the E. I. du Pont de Nemours Company.

3 October 2001; accepted 9 November 2001

Genome Sequence of the Plant Pathogen and Biotechnology Agent *Agrobacterium tumefaciens* C58

Brad Goodner,[1,2] Gregory Hinkle,[3] Stacie Gattung,[4] Nancy Miller,[4] Mary Blanchard,[4] Barbara Qurollo,[3] Barry S. Goldman,[3,4] Yongwei Cao,[3] Manor Askenazi,[3] Conrad Halling,[3] Lori Mullin,[3] Kathryn Houmiel,[4] Jeffrey Gordon,[3] Mark Vaudin,[4] Oleg Iartchouk,[3] Andrew Epp,[3] Fang Liu,[3] Clifford Wollam,[4] Mike Allinger,[4] Dahlia Doughty,[2] Charlaine Scott,[2] Courtney Lappas,[2] Brian Markelz,[2] Casey Flanagan,[2] Chris Crowell,[2] Jordan Gurson,[2] Caroline Lomo,[2] Carolyn Sear,[2] Graham Strub,[2] Chris Cielo,[2] Steven Slater[3]*

Agrobacterium tumefaciens is a plant pathogen capable of transferring a defined segment of DNA to a host plant, generating a gall tumor. Replacing the transferred tumor-inducing genes with exogenous DNA allows the introduction of any desired gene into the plant. Thus, *A. tumefaciens* has been critical for the development of modern plant genetics and agricultural biotechnology. Here we describe the genome of *A. tumefaciens* strain C58, which has an unusual structure consisting of one circular and one linear chromosome. We discuss genome architecture and evolution and additional genes potentially involved in virulence and metabolic parasitism of host plants.

Agrobacterium tumefaciens is a plant pathogen with the unique ability to transfer a defined segment of DNA to eukaryotes, where it integrates into the eukaryotic genome. This ability to transfer and integrate DNA is used for random mutagenesis and has been adapted into a powerful tool for production of transgenic plants, including soybean, maize, and cotton (*1*, *2*). *A. tumefaciens* was identified early in the 20th century as the causal agent of crown gall disease in plants (*3*). Pathogenesis is initiated when *Agrobacterium* detects small molecules released by actively growing cells in a plant wound. These molecules induce a series of virulence (*vir*) genes whose encoded products export the single-stranded transferred DNA (T-DNA) to the plant cell, where it integrates into the genome at an essentially random location. Once integrated, T-DNA gene expression alters plant hormone levels, leading to cell proliferation typical of a gall tumor. The T-DNA also encodes enzymes for the synthesis of opines, a class of nutrient molecules used almost exclusively by *A. tumefaciens* (*4–7*).

A. tumefaciens strains fall into three biovars, which differ in their host range, metabolic characteristics, relationships with other genera in the family Rhizobiaceae, and potentially their chromosome structure (*4–13*). The taxonomy of the Rhizobiaceae family is not without controversy, but we expect that

[1]Department of Biology, Hiram College, Hiram, OH 44234, USA. [2]Department of Biology, University of Richmond, Richmond, VA 23173, USA. [3]Cereon Genomics, LLC, 45 Sidney Street, Cambridge, MA 02139, USA. [4]Monsanto Company, 800 North Lindbergh Boulevard, St. Louis, MO 63167, USA.

*To whom correspondence should be addressed. E-mail: steven.c.slater@cereon.com